University of Alberta

# Is Syntax Intrinsic to Physics?

## An Examination of John Searle's Attack on Cognitivism

by

Darren Domsky ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master Of Arts

Department of Philosophy

Edmonton, Alberta

Fall 1998

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

*Your file Votre référence*

*Our file Notre référence*

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-34307-3

Canada

University of Alberta

Library Release Form

Name of Author: Darren Kirk Domsky

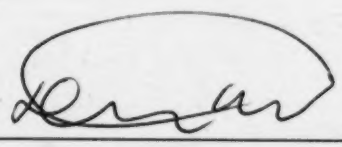
Title of Thesis: Is Syntax Intrinsic to Physics? An Examination of John Searle's Attack on Cognitivism

Degree: Master of Arts

Year this Degree Granted: 1998

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly, or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.



Aug 6, 1998

14209 24a Street
Edmonton, Alberta, Canada
T5Y 1N7

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled *Is Syntax Intrinsic to Physics? An Examination of John Searle's Attack on Cognitivism* in partial fulfillment of the requirements for the degree of Master of Arts.

Dr. Wesley E. Cooper

Dr. Martin Tweedale

Dr. Bruce A. Hunter

August 6, 1998

Dr. Oliver Schulte

Dr. Peter Dixon

August 6, 1998

**ABSTRACT**

John Searle attacks cognitivism, the belief that the brain is a digital computer, by arguing that, because computers are defined syntactically, and because syntax by definition is not intrinsic to physics but instead observer-relative, nothing, not even the brain, can be an intrinsic computer. He offers three avenues of support for his case, one based on the universal realization of computers, one on the invocation of tacit homunculi, and one on the similarity with non-natural function. I systematically undermine each of these avenues, and argue that he can give us no convincing reason to think that all syntax is observer-relative. I then show how that assumption is even potentially problematic, and conclude that Searle's attack on cognitivism is simply ungrounded. Lastly, I explore the concept of natural syntax, and suggest that at least some syntax might best be understood as a natural, observer-independent phenomenon.

**TABLE OF CONTENTS**


Chapter 1: Introduction.....1

Beyond the Chinese Room.....1

Searle's Arguments.....6

Chapter 2: The Universal Realization Argument.....8

The Universal Realizability of Computers.....8

Searle's Opposition.....14

A Proper Characterization.....25

An Assessment of the Argument.....28

Chapter 3: The Homunculus Argument.....33

The Homunculus Fallacy.....33

Homuncular Functionalism.....37

Searle's Homunculi.....40

Searle's Evidence.....42

Chapter 4: The Functions Argument.....46

Searle's View of Functions.....46

The Naturalistic View of Functions.....51

Searle's Attack on the Etiological Theory.....53

The Strength of Searle's Attack.....55

Chapter 5: The Nature of Syntax.....66

Intrinsic to Physics.....67

Slippery Slope.....74

All that is Required.....83

My Alternative Account.....84

A Final Assessment.....94

Bibliography.....96

# CHAPTER ONE: INTRODUCTION

## Beyond the Chinese Room

John Searle's attack on the computational theory of the mind originally began with the Chinese room argument.[1] The argument was intended to show that Strong Artificial Intelligence, the view that all there is to having a mind is having a program, was inevitably flawed. According to Searle, the argument demonstrated that, even if a system could follow a program so as to simulate human cognitive capacities perfectly, it would still not actually have those capacities, simply because the syntax of a program is never sufficient for the semantics of a mind. Searle's claim was that, even if a system running a program could perfectly simulate the ability to understand Chinese, to the point where it could pass any Turing Test[2] as a Chinese speaker, it would nevertheless have no understanding of Chinese.

The argument takes the form of a thought experiment. It asks us to imagine that someone who understands no Chinese sits in a room with a pile of Chinese symbols and a list of rules for transforming input to output by manipulating incoming symbols syntactically. When the room is asked a question in Chinese, it receives strings of Chinese symbols as input, and it answers

---

[1] Searle, 'Minds, Brains, and Programs'.

[2] The Turing Test, according to Simon Blackburn's Oxford Dictionary of Philosophy, is 'The test suggested by Alan Turing (1912-54) in his paper 'Computing Machinery and Intelligence' (*Mind*, 1950), for judging whether a machine is making an adequate simulation of the human mind. The test is also known as the imitation game. A person and the machine communicate with an interrogator who is kept apart. The interrogator may ask questions with the intent of distinguishing the human from the machine. The machine passes the test or wins the game if after an interval of time the interrogator cannot

those questions by producing strings of Chinese symbols as output. All the while, the person inside the room is manipulating symbols syntactically, completely ignorant of anything else. According to Searle, even if the list of rules being followed is complex enough to make the room's answers indistinguishable from a native Chinese speaker, neither the person inside nor any other part of the system literally understands Chinese. The syntactically defined list of rules is simply incapable of creating any kind of understanding. Because the programmed computer has nothing in addition to what the Chinese room has, Searle argues that the programmed computer cannot understand Chinese either. The Chinese room argument attacks Strong AI by showing that the syntax of a computer program is insufficient for the semantics of a mind.

In the ninth chapter of The Rediscovery of the Mind, Searle suddenly becomes more ambitious. In the chapter, aptly entitled "The Critique of Cognitive Reason", he goes beyond the Chinese room and attacks even weaker versions of artificial intelligence, views such as "that minds are partly constituted by unconscious computer programs".[3] Searle argues that even views such as these are mistaken, because 'cognitivism', the view that that the brain is a digital computer, is also fundamentally misguided. This is because syntax, according to which computation is defined, is not just insufficient for semantics, as the Chinese room showed, but entirely observer-relative as well. According to Searle, syntactical properties

---

distinguish the machine from the human."

"do not even exist except in the eyes of the beholder",[4] and are not discovered within physics but are instead assigned to physics. As Searle phrases it, "syntax is not intrinsic to physics."[5]

It is important to clarify this point. According to Searle, certain things in our world, such as gravitation and photosynthesis, are intrinsic features of reality because they are independent of any observer. Even if observers and users vanished from the planet, there would still be gravitation and photosynthesis. According to Searle, this is because ordinary features of reality always have observer-independent physical properties that determine what those features are. This is precisely why the natural sciences can discover and study them, because the physical properties are literally there to be discovered and studied. Symbols, on the other hand, have no physical properties that determine that they are symbols. In fact, the physics is wholly irrelevant except in so far as it allows an observer to assign symbols to it. Symbols, along with such things as chairs and bathtubs, are not intrinsic features of reality because they depend on observers to see them or interpret them as symbols, chairs, or bathtubs. For Searle, a symbol only exists if an observer assigns a symbolic interpretation to something. As a result, whereas ordinary physical properties are intrinsic to physics, syntax is not.

It should be noted that, for Searle, observer-relative does not mean arbitrary. He explains how we as observers might very

---

[3] Baumgartner, p.205

well have good, non-arbitrary reasons for assigning the observer-relative properties that we do. The point to remember is that, without observers, those properties would not exist in the first place. Searle explains:

> The assignment of observer-relative features to intrinsic features of the world is not arbitrary. Some intrinsic features of the world facilitate their use as chairs and bathtubs, for example. But the feature of being a chair or a bathtub or a nice day for a picnic is a feature that only exists relative to users and observers.[4]

Clearly, Searle does not want to equivocate observer-relative with arbitrary. That a certain property is dependent on observers for its existence does not mean that there are no non-arbitrary reasons why it is applied to the physical world.

It is because syntax is inherently observer relative that there is a serious problem with the concept of computation. Computation is defined in terms of symbol manipulation, but symbols, according to Searle, only exist when an observer assigns them. Therefore, computation only exists when an observer interprets some phenomenon computationally. This means that computation cannot be discovered in the world because computation does not even exist until an observer interprets it as computation. The brain could not be a computer because brains existed long before we as observers began to wonder about their composition. Cognitivism, as a result, is fundamentally doomed, because it is dedicated to the existence of something which cannot inherently exist in the natural world, namely a brain that is a digital computer.

---

[4] The Rediscovery of the Mind, p.215

Although Searle believes that the Chinese room argument is still valid, and that is still shows that the mind is not a computer program because the syntax of a program is not sufficient for the semantics of a mind, he now believes, in light of his new argument, that the Chinese room concedes too much. The computational theory of the mind is flawed not just because syntax is insufficient for semantics, but, more importantly, because syntax itself is not intrinsic to physics. Originally, Searle's conclusion was that the computational model was simply false, but now he concludes that it does not even reach the level of falsehood because it does not even make sense. In Searle's opinion, the question of whether or not something is a computer is utterly trivialized:

> The question "Is the brain a digital computer?" is ill defined. If it asks, "Can we assign a computational interpretation to the brain?" the answer is trivially yes, because we can assign a computational interpretation to anything. If it asks, "Are brain processes intrinsically computational?" the answer is trivially no, because nothing is intrinsically computational, except of course conscious agents intentionally going through computations.[7]

Before, Searle believed that the computational theory of the mind was logically possible but false; now, he sees it as unintelligible and incoherent.

Searle's attack on cognitivism can thus be seen as providing a further reason why we should accept the implications of the Chinese room argument. Strong artificial intelligence is doubly refuted, not just because it tries to reduce something with semantics to something that is insufficient for semantics,

---

[5] The Rediscovery of the Mind, p.208

but also because it tries to give an account of something intrinsic in terms of something observer-relative. According to Searle, the Chinese room, though effective, was never necessary; the new argument accomplishes all that the Chinese room did and more. The computational theory of the mind was already doomed, but now, in Searle's opinion, it is doomed twice over.

## Searle's Arguments

Searle's attack on cognitivism ultimately rests on the crucial premise that syntax is not intrinsic to physics. It is solely because syntax is an observer relative feature of the world and not an intrinsic feature of the world that the brain cannot be a syntactically defined computer. Because Searle takes this underlying premise to be self-evident and is therefore not especially concerned about defending it, piecing together coherent arguments from Searle's writing that support it is especially difficult. Nevertheless, as I see it, Searle can be interpreted as relying on three separate arguments to support his assumption that syntax is not intrinsic to physics. According to the first, the fact that syntax is not intrinsic to physics is made evident by the fact that the multiple realizability of computers inevitably collapses into the universal realizability of computers. The second argument states that syntax cannot be intrinsic to physics because whenever someone tries to treat syntax as though it were intrinsic to physics he unavoidably commits the homunculus fallacy. The third argument is an argument from analogy, and

---

[6] The Rediscovery of the Mind, p.211

states that syntax is not intrinsic to physics but instead observer relative in the same way that functions are never intrinsic to physics but instead always observer relative. I shall term these the universal realization argument, the homunculus argument, and the analogy with function argument.

In this thesis I plan to carefully analyze each of these arguments. One by one, I will show that they fail to establish Searle's central premise, that syntax is not intrinsic to physics. Searle's attack on cognitivism, on the belief that the brain is a digital computer, will therefore be shown to be wholly unsupported. Following that, I will examine two potential problems with Searle's central premise, the first being that it might be inconsistent with his claims about the nature of consciousness, and the second being that it might have the consequence that much of the external world must be observer-relative along with syntax and function. I will then go on to put forward an alternate view of syntax and its relation to physics, one that interprets syntax and function as natural phenomena. In particular, I will identify syntax as a natural phenomenon consisting in discrete physical phenomena, and will differentiate it from the labelling syntax we use to label those phenomena. In the end, I will show that Searle's attack on cognitivism ultimately fails, not only because it is entirely unsupported and potentially problematic, but also because it relies on an interpretation of syntax for which there is at least one plausible alternative.

---

[7] The Rediscovery of the Mind, p.225

## CHAPTER TWO: THE UNIVERSAL REALIZATION ARGUMENT

John Searle's first defence of his central premise that syntax is not intrinsic to premise is based on his claim the multiple realizability of computers inevitably slips into the universal realizability of computers. When Searle discusses the multiple realization of computers, and when he argues that there is no way to prevent that multiple realization from becoming the universal realizability of computers, he is very clearly providing evidence for his more central claim, that the syntax of a computer program cannot be intrinsic to the physics of a given system, but instead must be observer relative. When he explains that "universal realizability is ... a consequence of a much deeper point, namely that 'syntax' is not the name of a physical feature",[8] it becomes clear that his intention in even discussing the universal realizability of computers is to more clearly establish the fact that syntax is not intrinsic to physics. It is to the argument itself that we now turn.[9]

### The Universal Realizability of Computers

As Searle explains, it is routinely accepted that computers are, unproblematically, multiply realizable. It is generally agreed that computation can be realized in a variety of ways

---

[8] The Rediscovery of the Mind, p.209

[9] Searle, midway through presenting this argument, makes a curious disclaimer: "I do not think that the problem of universal realizability is a serious one. I think it is possible to block the result of universal realizability by tightening up our definition of computation." (The Rediscovery of the Mind, p.209) Shortly after, however, he disregards that disclaimer, saying that any such tightening up would be of "no help in the present discussion *because the really deep problem is that syntax is essentially an observer-relative notion.*" (Ibid.) Even though Searle labels the argument as not serious, he nevertheless presents it, and, to be as charitable as possible, I will examine the universal realizability argument as a potential support for Searle's central claim that

using a variety of materials. Voltage levels, cogs and levers, or cats, mice, and cheese might all conceivably be put together in such a way as to form a computer. He tells us that so long as a system admits of an interpretation consisting of 0's and 1's, that system, according to the philosophical community, can genuinely be said to be a computer that represents binary symbols. Just like thermostats and carburettors, computers can be constructed in a variety of functionally isomorphic ways.

Searle sees a strong difference, however, between the class of computers and the class of thermostats or carburettors. He claims that, unlike computers, thermostats and carburettors are "defined in terms of the production of certain *physical* effects."[10] If you want to find out if something is a thermostat or carburettor, you simply check if that something produces the appropriate physical effects. Computers, on the other hand, are "defined syntactically in terms of the *assignment* of 0's and 1's."[11] There simply are no physical effects of any kind that you can search for to determine if something is a computer or not. According to Searle, "To find out if an object is really a digital computer, it turns out that we do not actually have to look for 0's and 1's, etc.; rather we just have to look for something that we could *treat as* or *count as* or that *could be used to* function as 0's and 1's."[12] In other words, computers are not defined according to physical properties, but according to their potential for us as observers to assign certain

---

syntax is not intrinsic to physics.
[10] The Rediscovery of the Mind, p.207
[11] The Rediscovery of the Mind, p.207
[12] The Rediscovery of the Mind, p.206

syntactical interpretations to the physical properties involved. Unlike ordinary natural kinds, the class of computers cannot be identified by any sort of physical effect. Searle goes so far as to say that "The physics is irrelevant except in so far as it admits of the assignments of 0's and 1's and of state transitions between them."[13] A computer is not a computer because of any intrinsic physical feature, but rather because an observer decides to attribute a computational syntax to its physical features.

The reason, then, why computers are multiply realizable is entirely different from the reason why ordinary physical objects are multiply realizable. Carburettors can be realized in many ways because there are many physical ways that a carburettor could be assembled. The aim is defined physically, and the physics involved allows for multiple ways of achieving that aim. Computers, on the other hand, can be realized in many ways only because computers are completely syntactical entities, with no essential physical characteristics. All that is required for a computer to exist is an agent to believe that some physical system constitutes a computer. Because "*notions such as computation, algorithm, and program do not name intrinsic physical features of systems*",[14] the only thing, according to Searle, that makes a computer a computer is the assignment of its computational syntax by a conscious observer.

What is more, according to Searle, the observer is wholly unconstrained in his decision about what to treat as a computer.

---

[13] The Rediscovery of the Mind, p.207

Not only is it up to an observer to decide whether something is a computer in the first place, but the observer's decision is also entirely his to make. In Searle's view, because one can "describe anything in terms of 0's and 1's",[15] anything and everything can be interpreted as a computer, regardless of the physics involved. Any physical system is potentially a computer, and it actually becomes a computer the moment an observer decides to interpret it as such. Furthermore, so long as the system in question is a "sufficiently complex object",[16] Searle claims that there is some description of that object such that it is implementing any given program. Besides sufficient complexity, Searle sees no physical constraint on the number of possible programs that might be running in a system, leaving us with the result that not only is any particular physical system potentially running any given computer program, but that any given system is potentially running a very large number of computer programs simultaneously. Searle is careful not to say that the attribution of syntax is outright arbitrary. He believes that we often have very good reasons to assign the observer-relative properties that we do.[17] Still, Searle clearly believes that it is only the observer's whim that restricts his attribution of syntax. Searle's own examples attest to this: the "stomach, liver, heart, solar system, and the state of Kansas"[18] all qualify as potential computers, and the wall behind

---

[14] The Rediscovery of the Mind, p.210
[15] The Rediscovery of the Mind, p.208
[16] The Rediscovery of the Mind, p.208
[17] The Rediscovery of the Mind, p.211
[18] The Rediscovery of the Mind, p.208

Searle's back continually implements the Wordstar program.[19] So long as the required physical complexity is there, Searle believes that any physical system is potentially running any computer program, leaving the observer's attribution of syntax almost wholly unrestricted.

Because the observer has so much unconstrained freedom in assigning computation to the physical world, and because that assignment is all that is essential to computation, Searle claims that the multiple realizability of computers inevitably gives way to the universal realizability of computers. Everything in the world becomes a potential computer, and calling something a computer becomes an uninformative action. In Searle's view, the fact that there is nothing to prevent an observer from seeing literally any physical system as a computer ultimately proves that the actual syntax of computation is completely independent of the physics involved, and dependent only on the observer who imposes it on a physical system. That the multiple realization of computation inevitably slips into the universal realization of computers is therefore evidence that syntax is not intrinsic to physical systems at all, but instead is observer-dependent. Because the existence of a computational syntax is wholly dependant on the ascriptions of an observer, and because the physics involved is almost entirely irrelevant to the syntax that an observer might assign to it, syntax remains completely observer relative. The argument from universal realizability, in establishing that computers are not

[19] The Rediscovery of the Mind, p.208-209

just multiply but universally realizable, ultimately establishes the more general point that syntax is not intrinsic to physics.

The argument given by Searle to attack the computational theory of the mind is very similar to one given by Ian Hinckfuss to attack the functionalist theory of mind. Hinckfuss, at the Australasian Association of Philosophy Conference of 1978 at Canberra, presented the problematic case now referred to as 'Hinck's pail'. As described by Bill Lycan, it is very similar to Searle's argument:

> Suppose a transparent plastic pail of spring water is sitting in the sun. At the micro level, a vast seething complexity of things are going on: convection currents, frantic breeding of bacteria and other minuscule life forms, and so on. These things in turn require even more frantic activity at the molecular level to sustain them. Now is all this activity not complex enough that, simply by chance, it might realize a human program for a brief period (given suitable correlations between certain micro-events and the requisite input-, output-, and state-symbols of the program)? And if so, must the functionalist not conclude that the water in the pail briefly constitutes the body of a conscious being, and has thoughts and feelings and so on? Indeed, virtually any physical object under any conditions has enough activity going on within it at the molecular level that, if Hinckfuss is right about the pail of water, the functionalist quickly slips into panpsychism ...[20]

Just as Searle argues for the universal realizability and therefore vacuity of formally defined machines, Hinckfuss argues for the universal, or at least nearly universal, realizability and therefore vacuity of functionally defined machines. Searle's Wordstar Wall, as a problematic case for the computational theory of the mind, is therefore accompanied by the problematic case of Hinck's Pail.

---

[20] Lycan (1981), p.39

## Searle's Opposition

Not everyone, however, agrees with Searle that computation is universally realizable. David Chalmers, Ronald Endicott, Ned Block, Jack Copeland, and Ronald Chrisley have all objected that there are meaningful constraints on interpretation for computation, making the claim that something is a computer both informative and coherent. Although each philosopher objects to Searle's account of the universal realizability of computers in a different way, they all defend the coherence of the computational theory of the mind. In their opinion, there is good reason to believe that "there are additional constraints at work within our interpretive practice beyond the mere assignment of syntax that will isolate a limited set of physical systems and count only those systems as genuinely computational devices."[21] I will now summarize the various constraints on the interpretation of computation put forward by these philosophers.

One constraint on interpretation put forward by Chalmers, Endicott, Block, and Chrisley, and also one anticipated by Searle,[22] is that a system must exhibit appropriate causal relations between program states in order to truly be a computer. More simply, in order for computation to exist, "computational states must be able to cause other computational states to come about."[23] What is required, it is argued, is that the physical states that correspond, in an interpretation function, to certain computational states actually produce as a causal effect physical states that correspond to the appropriate

---

[21] Endicott, p.103-104

successor computational states. It is not enough to have a physical system that neatly maps out a series of physical states that correspond to computational states; the physical states must form a causal chain. As Endicott explains, Searle's discussion is limited to the concept of implemental capacity, which is "characterized by the mathematical notion of isomorphism whereby the physical states of the system are isomorphic to the transitions between states that the program specifies".[24] Actual implementation, on the other hand, is "determined by the appropriate causal interactions between the physical store of the program and the physical device itself."[25] Endicott argues that although a system may have a very large, and perhaps indefinite, number of implemental capacities, it has few actual causal implementations, and so computers are not universally but only multiply realized. Endicott summarizes beautifully:

> a system is a genuine computational device when there is a correspondence between its physical states and its formal states such that the causal structure of the physical system is isomorphic to the formal structure of the computational operations.[26]

Chalmers also believes that computers clearly require a causal structure. In his description of the formal criterion for the implementation of a combinatorial-state automaton, causality is central:

> A physical system P implements a CSA *M* if there is a decomposition of internal states of *P* into components $[s^1,\ldots,s^n]$, and a mapping *f* from the substates $s^i$ into

---

[22] The Rediscovery of the Mind, p.209
[23] Chrisley, p.67
[24] Endicott, p.104, footnote 6
[25] Endicott, p.104, footnote 6
[26] Endicott, p.104

> corresponding substates $S^i$ of $M$, along with similar decompositions and mappings for inputs and outputs, such that for every state transition rule $([I^1,\ldots,I^k], [S^1,\ldots,S^n]) \dashrightarrow ([S'^1,\ldots S'^n], [O^1,\ldots,O^l])$ of $M$: if $P$ is in internal state $[S^1,\ldots,S^n]$ and $[I^1,\ldots,I^k]$ respectively, this reliably causes it to enter an internal state and produce an output that map to $[S'^1,\ldots S'^n]$ and $[O^1,\ldots,O^l]$ respectively.[27]

According to Chalmers, causation is pivotal to the very concept of computation, for, in order for a computer to exist, it is required that certain internal states reliably cause other internal states within a physical system. One constraint, then, that Searle's critics offer on the interpretation of a system as a computer is that the system exhibit an appropriate causal structure.

Hilary Putnam has argued that the notion of computation does in fact need to incorporate causation. Unlike Searle, Putnam believes that a computational interpretation has to be a causal one if it is to properly characterize computation. Despite this admission, Putnam argues that computation is nevertheless a vacuous concept because it still follows that "Every ordinary open system is a realization of every abstract finite automaton",[28] and therefore that computation is still universally realized even though it is causal. Putnam adopts the notion of causation "that commonly obtains in mathematical physics,"[29] where the phrase "A causes B within system X" is interpreted to mean that

> any maximal state of the system which lies in the region of phase space corresponding to state A and which is compatible with the given boundary conditions

[27] Chalmers, p.318
[28] Putnam, p.121
[29] Putnam, p.96

> and with physical law will be followed by a maximal state which lies in the region of phase space corresponding to state B.[30]

Putnam's understanding of causation is one in which an omniscient being, such as Laplace's demon, "could *predict* that the system X would go into state B at the relevant time given the information that it was in state A at the earlier time and given the boundary conditions."[31] Putnam's account of causation incorporates the logical structure of material implication, where it is known that if the antecedent is true the conclusion is true and if the conclusion is false the antecedent is false. This is made clear by his own example:

> If, for example, under the given boundary conditions, a system has two possible trajectories--one in which Smith drops a stone on a glass *and* his face twitches at the same moment, and one in which he does not drop the stone and his face does not twitch--then "Mathematically Omniscient Jones" can predict, from just the boundary conditions and the law of the system, that *if* Smith (the glass breaker) twitches at time $t_0$, *then* the glass breaks at time $t_1$; and this relation is not distinguished ... from the relation between Smith's dropping the stone at $t_0$ and the glass breaking at $t_1$.[32]

Whereas Searle, in claiming that computation is universally realizable, insists that computation is non-causal, Putnam argues that computation is still universally realizable even though it does in fact have a causal structure.

According to Block and Chrisley, however, Putnam provides no assistance to Searle's attack on the computational theory of the mind because Putnam's understanding of causation, in their view, is intolerable. They argue not only that the concept of

---

[30] Putnam, p.96
[31] Putnam, p.96

computation must incorporate causality, but more specifically that it must incorporate a modal type of causality that supports counter-factuals over possible worlds differing minimally from the actual world. Under a modal account of causality, when we say that A caused B, we mean by this that if A had not been the case, then B would not have occurred. This understanding of causation, it is argued, is essential to the description of a computer, because it takes into account not only what a computer happens to do but also what it might have done instead, were the initial conditions different. As Block explains, "the isomorphism that makes a syntactical engine drive a semantic engine ... has to include not just a particular computation that the machine *does perform*, but also all the computations that the machine *could have* performed."[33] We need, in defining a computer, not only a sense that computational states cause other computational states, but also a sense that which states are caused now strongly depends on which states existed prior. According to Block, Searle's wall cannot be a computer because the wall simply does not have the right kind of modal causal structure. For clarification, he gives the example of an X-OR gate:

> In order for the wall to be this computer, it isn't enough for it to have states that correspond to '0' and '1' followed by a state that corresponds to '1'. It must also be such that *had* the '1' input been replaced by a '0' input, the '1' output *would have been* replaced by the '0' output. In other words, it has to have symbolic states that satisfy not only the *actual* computation, but also the *possible* computations

[32] Putnam, p.97
[33] Block, p.399

> that the computer *could* have performed. And this is nontrivial.[34]

According to Chrisley, it is the modal nature of the causality of computation that prevents the multiple realizability from slipping into the universal realizability of computers. When someone tries to arbitrarily map a computational syntax onto something like a wall, that person "is appealing to fortuitous patterns in the physical events, and not patterns that are preserved over variations in possible worlds."[35] For Searle, if syntactical unit A happens to precede syntactical unit B, then we have an instantiation of computation, and for Putnam, if causal event B happens to succeed causal event A, then we have causality. Chrisley believes we need more than this. B cannot happen to follow A; it needs to always follow A in some range of possible worlds.[36] The constraint on the interpretation of a computer provided by Block and Chrisley is that computation must be not only causal but modal as well.

Copeland offers a further constraint on the interpretation of computers, one that strongly supports and deepens this counter-factual understanding of computation. He argues that, if we are to properly assign computation to physical systems, the truth condition assigned to each computational axiom cannot be a universally quantified material implication. Instead, the if-then relationship between computational states must be the strong dependency relationship of the counter-factual conditional, where "if A then B" is interpreted to mean "if A

---

[34] Block, p.399-400
[35] Chrisley, p.73
[36] Chrisley, p.73

were to happen B would", or "if A were to have happened then B would have happened".[37] Since any material implication with a false antecedent or a true consequent immediately becomes true, it is far too easy for the relation to be satisfied, and, as a result, far too many physical systems become computers, rendering the notion of computation vacuous. We require, Copeland argues, a strong dependency relationship of the kind expressed by Alan Turing's original phrase 'completely determined by'. As Turing himself explained, "If at each state the motion of a machine ... is *completely* determined by the configuration, we shall call the machine an 'automatic machine'."[38] Because the lenient concept of material implication allows far too many if-then statements to be realized and therefore too many physical systems to be computers, another constraint offered by Searle's critics on the interpretation of a computer is that strong conditionals rather than material implications be applied to computational states.

Another restriction on the notion of computation offered by Copeland is that computers must be capable of acting within the world. Copeland points out how Searle's explanation of why his wall is a computer running the Wordstar program is extremely unsatisfactory because the dynamic logics of a computer are "logics of computer *action*."[39] According to Copeland, Searle's Wordstar wall is not a real computer because

> all the computational activity occurred *outside* the wall, in the course of obtaining Table I, which is simply a record of the activity within the cpu of the

[37] Blackburn, p.86
[38] Turing, (1936), p.231-232
[39] Copeland, p.348

> machine that actually performed the computation. Once Table I is secured the labelling scheme is constructed from it ex post facto. The wall under this novel description is at most a passive 'scoreboard' and is no more an active participant in the computation than the scoreboard is an active player in a game of billiards.[40]

The counter-intuitive aspect of Searle's computational wall is that all of the active computation is performed by Searle and not the wall. Copeland argues that the concept of computation is a concept of action rather than of interpretation, whereby computers actually compute. They *act* in accordance with the instructions of the algorithm of a program, which is why we use the active voice when we say that a computer executes a program. When Searle attributes computation to a wall, Copeland points out that the active component is missing, and we find that all of the action takes place in the formation of a table by an observer. Whereas our intuition seems to be that computers, in order to run a program, must actively compute, Searle requires that computation be attributed ex post facto by an external observer. Copeland, then, would insist that a computer, in order to exist, must actively compute rather than act as a passive locus for computational interpretation.

Yet another constraint on the interpretation of a computer, as offered by both Copeland and Chrisley, is that computers must be generative, or non-temporally specified. As Chrisley explains, interpretations like Searle's Wordstar interpretation "are not generative: They do not determine a means for mapping physical states to computational states outside of the time

---

[40] Copeland, p.348

window for which he actually has observations of the system."[41] Searle's understanding of computation, Chrisley explains, incorporates a problematic temporal specificity. Computers become objects fixed in time, with computational lifespans which begin and end at very specific times. As Copeland explains,

> The labelling scheme used in the model is, of course, incomplete in that the scheme identifies the regions of the wall that bear labels at the designated moments $t_1, \ldots, t_n$ but provides no information concerning which regions of the wall are the label-bearers at times prior to $t_1$ or subsequent to $t_n$.[42]

Clearly, a Searlean computer, with its fixed time frame, is vastly unlike the everyday notion of a computer, whose "labelling scheme remains applicable throughout the lifetime of the entity".[43] Copeland continues:

> The axioms as interpreted have no entailments concerning times lying outside this range, whereas under their intended interpretation the axioms are (and entail) conditional statements that are true at any moment during the normal functioning of the device. (Is the wall perhaps in a state of malfunction at all time prior to $t_1$ and subsequent to $t_n$? Hardly. At the whim of the modeller the wall can be made to run the program again, say through the moments $t_{n+1}, \ldots, t_{2n}$.)[44]

According to Copeland, computers, on Searle's understanding of them, become very counter-intuitive, temporally-fixed things. What is perhaps even more counter-intuitive is that, as a result of this temporal-fixedness, the very existence of a computer becomes necessarily past-tense. Searle's theory of computation implies that statements about computation are of necessity statements about the past because the act of attributing a

---

[41] Chrisley, p.74
[42] Copeland, p.348-349
[43] Copeland, p.349
[44] Copeland, p.349

computational syntax to a physical system is always a reflexive process. Under Searle's framework, computer actions are restricted not only to specific time frames but also to the past. Therefore, it follows that Searlean computers can never presently exist; they can only be attributed to finite and past time intervals. According to Copeland, this is in direct conflict with our intuition that "to describe a physical entity as a computing machine of a certain kind is to envisage being able to *predict* aspects of its physical behaviour on the basis of its architecture-algorithm specification and its labelling scheme."[45] Clearly, there is no room for the prediction of the behaviour of a Searlean computer, since it cannot even exist, let alone behave, in the present. A final constraint, then, on the interpretation of a computer as offered by Copeland and Chrisley is that computers must be generative, and not fixed in time or necessarily located in the past.

When we look at the various additional constraints on interpretation provided by these philosophers, it becomes clear that they believe that Searle radically mischaracterizes and underdefines computation. In their opinion, Searle simply does not appreciate how difficult it is for computation to be instantiated. They claim that, because he leaves out crucial criteria, Searle provides an overly simplistic and naive characterization of computers. Searle, in his own words, demonstrates how simple his one criterion for the existence of computation really is:

---

[45] Copeland, p.349

> We were looking for some facts of the matter which would make the brain processes computational; but given the way we have defined computation, there never could be any such facts of the matter. We can't, on the one hand, say that anything is a digital computer if we can assign a syntax to it, and then suppose there is a factual question intrinsic to its physical operation whether or not a natural system such as the brain is a digital computer.[46]

As Endicott explains, this simply is not an accurate description of how computation is defined:

> For the cognitivist does *not* say 'anything is a digital computer if we can assign a syntax to it,' pure and simple, and then go on to 'suppose there is a factual question intrinsic to its physical operation' about whether the brain is a computer. Rather, the cognitivist says that anything is a computer *if we can assign a syntax to it under the conditions specified by the additional constraints on interpretation, ...* in light of which there *is* a fact of the matter which distinguishes the class of things that are intrinsically computers from Searle's much wider set of universally realized and observer-relative computers.[47]

The essential problem with Searle's account of computation, according to these five philosophers, is that he defines computation too simply. Searle holds that the only thing that is essential to computation is that an observer make a computational interpretation. In his opinion, it is the assignment of a computational syntax to a system and *only* the assignment of that syntax which brings a computer into existence. According to Searle's critics, this is false. They claim that something is a computer if a computational syntax can be accurately attributed to it under the additional restrictions that they summarize. As a result, the universal realizability

---

[46] The Rediscovery of the Mind, p.209-210
[47] Endicott, p.106

of computers fizzles, and we are left with only a multiply realizable phenomenon.

Computation, according to Searle's critics, in not only just multiply realizable but also extremely rare. When the constraints summarized are applied to our syntactical ascriptions, it is clear that very few physical systems actually qualify as genuine computers running actual programs. According to Chalmers' calculations, "If we take a CSA whose state vectors have one thousand elements, with ten possibilities for each element, then ... the chance of an arbitrary set of physical states having the requisite causal relations is something less than one in [$10^{1000}$ to the power of $10^{1000}$]".[48] If we incorporate the restrictions on interpretation, the claim that Searle's wall is actually running the Wordstar program becomes absurd. Problem cases such as Searle's wall or Hinck's pail quickly dissolve as problem cases, for the odds of such systems ever actually instantiating a specific computation are ludicrously poor.

## A Proper Characterization

How, then, should we understand Searle's account of computation? The general consensus among the five philosophers under review is that, although Searle's claim that anything and everything is a potential computer is false, his underlying argument does have some merit. They admit that to an extent Searle is on the right track, for the observer does play a part in determining what is a computer and what is not. Chalmers

[48] Chalmers, p.319

admits that, with computational interpretations, "there is a limited degree of observer relativity: any given physical system will implement a number of computations, and which one of these an observer chooses to focus on will depend on her purposes",[49] and Endicott accepts that syntax is "observer relative with respect to an assignment of zeros and ones to physical variables *tout court*".[50] However, as Block explains, even though "We have a great deal of freedom as to how to interpret a device, ... there are also very important restrictions on this freedom".[51] In other words, although "Searle is right that whether something is a computer and what computer it is is *in part* up to us ... it is not *totally* up to us."[52] Thus, even though the observer does play a role in determining what syntax is instantiated in a given physical system, the idea put forward by Searle's critics is that this role is a severely limited one. They claim that, ultimately, even though computational interpretations are in part relative to the observer, the possible interpretations open to that observer are strongly restricted by the constraints already mentioned. It is not claimed, then, that Searle is completely mistaken in his belief that computation is not multiply but universally realizable. Rather, the understanding these five philosophers seem to have is that Searle takes a valid point too far, by focusing on the observer-relative component alone, and by not taking seriously any of the additional constraints that might limit computation.

---

[49] Chalmers, p.319
[50] Endicott, p.107
[51] Block, p.400
[52] Block, p.400

Copeland in particular has an interesting interpretation of Searle's universal realization argument. He goes to great lengths to defend what he describes as *Searle's Theorem*, which is summarized as follows:

> For any entity $x$ (with a sufficiently large number of discriminable parts) and for any architecture-algorithm specification $y$ there exists a labelling scheme $L$ such that $(x,L)$ is a model of $y$.[53]

Copeland takes the time to demonstrate that Searle is technically correct when he claims that "there is some pattern of molecule movements that is isomorphic with the formal structure of Wordstar",[54] but then goes on to explain that Searle is correct only on a nonstandard interpretation of computational theory. By a nonstandard interpretation, Copeland means an interpretation "that does not respect the intended meanings of the terms of the theory."[55] Copeland explains how an honest model of a computational algorithm is one that uses the relevant terms the way they were originally intended. In the case of computation, this means that when Searle ascribes computational syntaxes without regard for the additional constraints on interpretation already outlined his ascription is no longer truly computational in an honest sense. In short, Copeland believes that Searle can in fact find a pattern in a sufficiently complex system and then attribute to it a syntax that will be isomorphic with the syntax of a computer running a given program, but he argues that this isomorphism is unimportant. He explains that "it would be an elementary error

---

[53] Copeland, p.339
[54] The Rediscovery of the Mind, p.209
[55] Copeland, p.346

to infer from this that the wall computes in any genuine sense",[56] and asserts that, unless a model of a computational algorithm is made honest by the observer respecting the intended meanings of the terms involved, the computational interpretation becomes valid but entirely uninteresting. Ultimately, Copeland believes that Searle is right in his claim that an observer can assign any computational syntax to any sufficiently complex system. The point is that such an assignment would simply be irrelevant, because the interpretation would not respect the intended meanings of the computational concepts. In the end, Copeland claims that Searle's argument from the universal realizability of computers fails, not because it is logically flawed, but because Searle's understanding of the ascription of computational syntax no longer has anything to do with computation. Without the additional constraints on interpretation already summarized, Searle's syntax is no longer computational, and therefore has no bearing on the nature of computation.

## An Assessment of the Argument

How does Searle's original argument fair against these harsh interpretations? Certainly, his argument from universal realizability argument cannot be outwardly dismissed. In the first place, one would think that if Searle's argument truly were fundamentally misguided there would be a strong degree of unanimity amongst his critics. This certainly is not the case. On the contrary, Searle's critics have a variety of positions on

[56] Copeland, p.348

the basic issues, and, if nothing else, Searle, in forming his argument, manages to point out how much disunity there is in the cognitivist camp. If, as Searle's critics claim, the ascription of computation truly is restricted by the constraints on interpretation already summarized, then one might ask why there is so much disagreement about when computation is actually realized. In fact, each of the five philosophers mentioned has a different understanding of what properly constrains a computational interpretation. For instance, Copeland, unlike the others, believes that computation need not have a causal structure. He claims that "a causal account of computation lacks the necessary generality",[57] and excludes abstract but genuine computers such as Turing machines. The idea that computation requires a modal causal structure is foreign to the writing of Endicott, Chalmers, and of course Copeland, and the belief that computation must employ the counter-factual conditional rather than material implication is held by no one but Copeland. Copeland alone believes that computers must act in order to truly compute, and only Copeland and Chrisley discuss the requirement that a computer be generative rather than fixed in time. If Searle's argument from the universal realization of computers really is in error, and if it therefore really does fail to establish or provide evidence for the claim that syntax is not intrinsic to physics, it seems odd that there would be so much disagreement among the defenders of cognitivism about the requirements of computation.

---

[57] Copeland, p.353

In addition to pointing out the disunity of his critics, Searle might also argue that the constraints on interpretation provided simply cannot be fairly applied to computation. For instance, with regard to the restriction that computation must have an appropriate causal structure, modal or otherwise, Searle clearly believes that a fundamental misunderstanding is occurring. In his opinion, syntax, no matter how it is ascribed onto a physical system, can never have any causal powers because syntax does not even exist independent of observers. Since syntax in his opinion must be ascribed by an observer to even exist, syntactically defined entities such as computers cannot be causal because the units of syntax on which computers are based "do not even exist except in the eyes of the beholder."[58] It would seem that Searle would reject the possibility that causality might constrain the realization of computation because the syntax of computation cannot exist independent of observers, and therefore cannot cause anything. As for such constraints as that computers must be generative or that computers must act, Searle would most likely protest again that such constraints cannot possibly apply to observer-relative entities. For instance, Copeland points out that all that action for Searle's Wordstar wall takes place in Searle, who observes the activities of his wall in such a way as to form an appropriate labelling scheme. Copeland suggests, as a response to this allegedly counter-intuitive fact, that computers should be defined as things which must act. Searle, however, would argue that this

[58] The Rediscovery of the Mind, p.215

constraint makes no sense. How, he might ask, could an observer-relative entity *do anything*? The requirement that computers must act, Searle would say, is problematic because computers, by their very nature, are always the passive objects of the intentional thoughts of conscious human beings. Searle would argue that of course all of the action in his Wordstar wall takes place inside him as the outside observer. When the object in question only exists relative to observers, how else, Searle might ask, could it be? Clearly, Searle would argue that the constraints on interpretation summarized do not apply to computation, and that his argument from universal realizability therefore holds.

Certainly, Searle's five critics cannot offer an infallible rebuttal to such Searlean counter-arguments. Battles of intuition are extremely difficult to settle, and, at this point in the discussion, the issue remains unsettled. Nevertheless, it is clear that at the very least Searle's universal realization argument cannot be used to support the broader claim that syntax is not intrinsic to physics. Ultimately, Searle cannot dismiss the various constraints on interpretation put forward against his universal realization argument without assuming precisely what the argument is attempting to establish, namely that syntax is observer-relative and therefore extrinsic to physics. In the end, Searle's universal realization argument is question-begging. In order to use the universal realization of computers as evidence of the fact that syntax is not intrinsic to physics, Searle must first assume that syntax is

not intrinsic to physics in order to establish that computers are universally realizable. Searle cannot convince us that computation is universally realizable unless we already believe that syntax is an observer-relative, non-intrinsic phenomenon. The universal realization argument is therefore fraught with circularity, and establishes nothing.

The issue of what actually constrains our interpretation of computers will therefore be set aside. Because it seems unclear which particular picture of computation one ought to endorse, the appropriate thing would seem to be to leave the matter unresolved. Certainly, more work needs to be done in finding the solution to this problem. For now, all that we can say is that at the very least there is as much reason to believe that our interpretation of computation is constrained as there is to believe that it is not. At the end of the day, Searle's universal realization argument fails to establish that syntax is not intrinsic to physics. Searle presents the would be fact that the multiple realization of computers inevitably slips into the universal realization of computers as "just the tip of a whole iceberg of problems."[59] In reality, however, it would seem that the argument does not establish any such problems without first assuming the truth of those problems to begin with. Thus, we will have to turn to Searle's second piece of evidence for his general claim that syntax is not intrinsic to physics.

---

[59] The Rediscovery of the Mind, p.207

## CHAPTER THREE: THE HOMUNCULUS ARGUMENT

In this chapter, I will examine Searle's second line of defence for his central claim that syntax is observer-relative and therefore not intrinsic to physics. His homunculus argument, as I will call it, is certainly never spelled out very explicitly, and some active participation by the reader is required if a clear argument is to be identified. Nevertheless, Searle very clearly puts it forward as a second argument for how we know that syntax is extrinsic to physics. Syntax must be extrinsic to physics, he argues, because whenever we take syntax to be intrinsic to physics we inescapably commit the homunculus fallacy. Whenever we try to treat something as an intrinsic computer, our mistake is made blatant by the way we consistently and necessarily invoke homunculi within that thing.

### The Homunculus Fallacy

Cognitive science, explains Searle, has long been infected with the homunculus fallacy. In the computational theory of the mind, he explains, "the idea always is to treat the brain as if there were some agent inside it using it to compute with."[60] Searle cites David Marr (1982) in particular, and recounts how Marr "describes the task of vision as proceeding from a two-dimensional visual array on the retina to a three-dimensional description of the external world as output of the visual system."[61] The problem, according to Searle, is that we are left

[60] The Rediscovery of the Mind, p.212
[61] The Rediscovery of the Mind, p.212

to wonder "who is reading the description."[62] It would seem, claims Searle, that in most works in the computational theory of the mind it is necessary to invoke a homunculus inside the computational system in order to treat its operations as intrinsically computational. In essence, whenever we try to treat something as an intrinsic computer, we unavoidably hatch homunculi and set them the task of being the observers who attribute computation to purely physical phenomenon.

Although Searle never expressly says why invoking homunculi is a bad thing, it is not difficult to fill in the blanks. In fact, there are two potential reasons why Searle might claim we should shun homunculi. Certainly, the invocation of homunculi carries with it the threat of a vicious infinite regress. If brains are explained in terms of other brains, then of course the explanation will never end, since each new homunculus requires its own additional homunculus. B. F. Skinner provides a witty anecdote stressing this point:

> ... the little man ... was recently the hero of a television program called "Gateways to the Mind" ... The viewer learned, from animated cartoons, that when a man's finger is pricked, electrical impulses resembling flashes of lightning run up the afferent nerves and appear on a television screen in the brain. The little man wakes up, sees the flashing screen, reaches out, and pulls the lever ... More flashes of lightning go down the nerves to the muscles, which then contract, as the finger is pulled away from the threatening stimulus. The behavior of the homunculus was, of course, not explained. An explanation would presumably require another film. And it, in turn, another.[63]

---

[62] The Rediscovery of the Mind, p.212
[63] Skinner, p.80

Another reason Searle might cite in his dislike for homunculi is that invoking homunculi begs cognitive-psychological questions. As Dennett explains, "since psychology's task is to account for the intelligence or rationality of men and animals, it cannot fulfil its task if anywhere along the line it *presupposes* intelligence or rationality."[64] Clearly, any explanation of intelligence in terms of intelligence is only question-begging, and provides no real account of intelligence at all. Regardless of which reason Searle would subscribe to, it is clear that the invocation of homunculi is potentially problematic.

Searle's point in talking about homunculi is not that computers cannot exist, but rather that intrinsic, userless computers cannot exist. He explains how, with regular desktop computers, "there is no homunculus problem, because each user is the homunculus in question."[65] Whenever we use a computer, we are the ones who attribute the syntax to the machine and interpret its actions as computational.[66] However, when something is supposed to be an intrinsic computer, "we are still faced with the question 'And who is the user?'"[67] The only way, according to Searle, that we can put a syntax into the physics of a system is to invoke a homunculus for that system, and,

---

[64] Dennett, (1978), p.58

[65] The Rediscovery of the Mind, p.214

[66] It has been brought to my attention by Wesley Cooper that Searle is not quite correct on this point. Although the individual user interprets what is on his computer screen as text rather than squiggles, most computer-users are completely ignorant of the programming syntax or the electrical-mechanical syntax of the machine's inner workings. It would be useful to differentiate between the user-homunculus, who interprets certain output syntactically, and the engineer-homunculus, who interprets certain voltage levels syntactically, and it would be more accurate for Searle to say that, for regular desktop computers, there is no homunculus problem because a loose community of computer engineers and machine language coders is the group-homunculus for a given machine. Since nothing depends on this point, however, I will let it drop.

[67] The Rediscovery of the Mind, p.214

because we are forced to take such an erroneous step, it becomes clear that nothing can be an intrinsic computer.

What is more, in Searle's opinion, there is no way to discharge the homunculus and thus rescue cognitivism. As he explains, many writers are undisturbed by the homunculus fallacy because they feel that the homunculus within the intrinsic computer can be discharged. Citing Dennett, (1978), Searle explains how many philosophers believe that a recursive decomposition of higher-level homunculi to progressively stupider homunculi and finally to simple flip-flops that do not involve a homunculi at all can neatly eliminate all the homunculi in question. The idea is to extinguish homunculi by replacing them with extremely simple physical operations which require no homunculi at all.

Searle's response to this line of thinking is to say that the reason why the homunculus is invoked in the first place ultimately prevents it from ever being discharged. The homunculus is required to attach the computational syntax onto the physics involved, and, as a result, discharging it becomes impossible. In Searle's words,

> we still must face the question we had before: What facts intrinsic to the system make it syntactical? What facts about the bottom level or any other level make these operations into 0's and 1's? *Without a homunculus that stands outside the recursive decomposition, we do not even have a syntax to operate with.* The attempt to eliminate the homunculus fallacy through recursive decomposition fails, because the only way to get the syntax intrinsic to the physics is to put a homunculus in the physics.[68]

[68] The Rediscovery of the Mind, p.213-214

Thus, syntax is shown to be extrinsic to physics because whenever syntax is treated as though it were intrinsic to physics, or, more specifically, whenever a purely physical system is treated as an intrinsic computer, a tacit homunculus is unavoidably invoked. The homunculus argument therefore demonstrates that syntax cannot be intrinsic to physics.

## Homuncular Functionalism

Homunculi have long been a regular part of the functional analysis of the brain. The idea dates back to Fodor's 1968 work,[69] and has become quite commonplace in functionalist theories of the mind. The general idea is fairly simple:

> Think of the human mind as represented by an intelligent being in the head, a "homunculus." This of this homunculus as composed of smaller and stupider homunculi, and each of these being composed of still smaller and still stupider homunculi until you reach a level of completely mechanical homunculi.[70]

Although homuncular functionalism takes its name from the often-ridiculed view that cognitive functions are performed by a little person inside the brain, it clearly distances itself from that view and the problems discussed earlier. The idea is not to replace the mind with smaller, stupider minds, but to explain the mind in terms of more and more precise functional capacities. Under homuncular functionalism,

> ... behavioural data are to be seen as manifestations of subjects' psychological capacities, and those capacities are to be explained by understanding the subjects as systems of interconnected components. Each component is a 'homunculus', in that it is thought of as a little agent or bureaucrat operating within its containing subject; it is identified by reference to the function it performs. And the

[69] Fodor, (1968), p.20
[70] Block, (1995), p.385

> various homuncular components co-operate with each other in such a way as to produce overall behavioural responses to stimuli.[71]

The reduction of complex capacities to simple capacities continues until a fundamental, bottom-level is reached, and the processes involved become so simple that they become primitive processes which can be executed by purely mechanical primitive processors. Very complex functional tasks are therefore broken down into simpler and simpler tasks until a large set of extremely simple tasks is all that remains. As a result, extremely complex and even intelligent capacities are understood in terms of very mechanically-defined capacities.

This concept of hierarchical control is fairly intuitive where ordinary physical systems are concerned. Understanding how an automobile functions, for instance, is clearly a matter of understanding how its various components and sub-components function together. Lycan explains:

> An automobile works--locomotes--by having a fuel reservoir, a fuel line, a carburettor, a combustion chamber, an ignition system, a transmission and wheels that turn. If one wants to know how the carburettor works, one will be told what its parts are and how they work together to infuse oxygen into fuel; and so on.[72]

What the homuncular functionalist wants to do is carry this idea of hierarchical control over to the field of cognitive science. For instance, according to homuncular functionalism, our capacity to understand language can be explained in terms of simpler and simpler capacities that ultimately bottom out in

---

[71] Lycan, (1996), p.176
[72] Lycan, (1996) p.178

purely mechanical brain processes. Ned Block explains how such an analysis of human language comprehension might work:

> Part of the system will recognize individual words. This word-recognizer might be composed of three components, one of which has the task of fetching each incoming word, one at a time, and passing it to a second component. The second component includes a dictionary, that is, a list of all the words in the vocabulary, together with syntactic and semantic information about each word. This second component compares the target word with words in the vocabulary (perhaps executing many such comparisons simultaneously) until it gets a match. When it finds a match, it sends a signal to a third component whose job it is to retrieve the syntactic and semantic information stored in the dictionary.[73]

According to homuncular functionalists, even very intelligent cognitive competencies can be explained by an appeal to simpler cognitive competencies, and accomplishments as impressive as language comprehension can ultimately be reduced to sets of unimpressive, completely mechanical accomplishments.

The homuncular functionalist is therefore not interested in how a specific homunculus works or even in how it is physically realized, but only in what it does. The word 'homunculus' is thus used extremely loosely. Electronic gates, whole computers, or even people could all potentially fill the role of a given homunculus. All that is required is that the homunculus physically instantiate the required input-output function. So long as the job is done and the relevant input-output function is effected, what actually fills the role of a given homunculus simply does not matter. The fact that potentially embarrassing homunculi are discussed at all only emphasizes the functionalist's disinterest in whatever thing fills a functional

role. As Dennett explains, "one says, in effect: put a little man in there to do the job."[74] The homuncular functionalist identifies the homunculus only by the task it carries out, and nothing more.

**Searle's Homunculi**

It is clear that the homunculus Searle has in mind when he claims that the homunculus fallacy is endemic to cognitivism is quite different from the one discussed by homuncular functionalists. Whereas the functionalists very clearly are dealing with a functional analysis of physical systems, and are trying to break complex functional capacities into more and more simple capacities, Searle is concerned with the ontology of observers, and is concerned only with whether or not an observer is being posited where there clearly is no, and can be no, observer. When Searle charges the cognitivist with invoking homunculi in order to treat the brain as an intrinsic computer, he is clearly not talking about functional capacities at all. Searle's intent is to reassociate cognitivism with the ridiculous view that homuncular functionalism takes its name from; namely, the view that cognitive functions are performed by some variation of demon within the brain. It is little observers, and not little task undertakers, that Searle wants to associate with the view that the brain in a computer program. Searle is arguing that the computational theory of the mind, in order to treat the brain as an intrinsic computer, unavoidably

---

[73] Block, (1995), p.385-386
[74] Dennett, (1978), p.124

and even comically requires the invocation of literal observers within the brain.

In light of Searle's account of the nature of the homunculus being invoked, his strong reaction to the possibility that the homunculus might be discharged is more understandable. For Searle, the reason that a homunculus must be posited in the first place is because an observer is required to fix a syntax onto a particular physical system. When homuncular functionalists start talking about replacing individual homunculi with simpler and simpler homunculi until the last homunculi are replaced with brute mechanical processes, Searle's natural response is to wonder how an observer can possibly be replaced by physical processes which clearly observe nothing. In his opinion, whether or not certain input-output functions are realized simply skirts the issue. The underlying reason why the homunculus is necessary in the first place is the same reason why the homunculus will not go away: the syntax-blind physics still requires an observer if the computational syntax is to exist.

What we have, then, are two different kinds of homunculi. Searle, in effect, is introducing an entirely new kind of homunculus, one that makes an observation rather than fills a functional role. It is this observer-homunculus, Searle claims, that brings cognitivism to its knees, for if we do not invoke it within the brain there can be no syntax within the brain, and if we do invoke it we end up in a ridiculous, irreparable position.

## Searle's Evidence

Does Searle have any compelling grounds for us to accept the claim that cognitivism unavoidably invokes problematic, observation making bogeymen? Certainly, this is the crucial question. If he can provide such grounds, then the crucial premise in his overall argument, that syntax is not intrinsic to physics, is secured, and his attack on cognitivism succeeds. Searle claims that if we try to treat a physical system as an intrinsic computer, we necessarily invoke homunculi, and not the kind that carry out functional tasks, but instead the kind that make observations. Can he give us good reason to agree with him?

Certainly, homuncular functionalists believe we have good reason to invoke functional-homunculi. According to Andrew Ward, we need to ask only one question:

> Does it help in making as much sense as possible of the intentional behavior of a sentient creature to posit homunculi, to who whom intentionally characterized abilities and activities are attributed? If so, then the representationalist will posit homunculi. ... what intentionally characterized abilities and activities are attributed to the posited homunculi will be constrained by the representationalist's goal of making as much sense as possible ... of the sentient creature's intentional behavior.[75]

The homuncular functionalist's claim is that functional-homunculi are posited if and when it is fruitful to do so. Homunculi are thus seen as useful devices for gaining a greater understanding of complex systems such as brains, and it is their usefulness that makes us interested in invoking them. As

---

[75] Ward, p.66

Dennett explains, when we invoke functional-homunculi, it is always because it is extremely useful. We posit homunculi because "there seems no better way of describing what is going on."[76] Certainly, the functionalist camp is confident that we have good reason for invoking functional-homunculi. Can Searle provide us with good reason why observer-homunculi are unavoidably invoked?

One potential avenue of support for Searle's claim might be the Chinese Room argument. The reason, it might be argued, why we need to posit a problematic homunculus within the brain in order to treat it as an intrinsic computer is because the brain clearly has intentionality and, as the Chinese Room shows, syntactically defined computers cannot have intentionality. On the grounds that syntax is insufficient for semantics, Searle might very well argue that if the brain really is a computer then it requires an intentional computer-user. I will not pursue this line of defense, however, for two reasons. To begin with, the Chinese Room argument has its own array of criticisms, and the debate over what if anything it proves is still very heated. To use the Chinese Room to defend the homunculus argument and ultimately the claim that syntax is not intrinsic to physics would only complicate the issue. More importantly, I will avoid taking the Chinese Room argument as evidence for the current claim because Searle explicitly says that the two arguments are unrelated and independent of each other. According to Searle, that syntax is not intrinsic to physics is

[76] Dennett, (1978), p.123

a "separate and different point".[77] Not only is the current argument different, but, in Searle's opinion, it is actually superior to the Chinese Room. As he explains, "for the purposes of the original argument, I was simply assuming that the syntactical characterization of the computer was unproblematic. But that is a mistake."[78] For these reasons, the Chinese Room argument will not be taken as evidence for the current claim.

What does that leave us? The only remaining thing that Searle can rest his homunculus argument on is the premise that an observer is required to attach a syntax to a given physical system. Where an intrinsic computer is concerned, human beings obviously cannot perform this role and so the invocation of an observer-homunculus becomes necessary. For Searle, the entire homunculus argument depends on the assumption that "the only way to get the syntax intrinsic to physics is to put a homunculus in the physics."[79] This assumption is also what prevents the homunculus from being discharged, since purely mechanical processes, no matter what functional tasks they might replace, cannot make observations and therefore cannot attach the required syntax to the physical system in question. When he asks "who is reading the description?",[80] Searle is clearly leaning on the assumption that, wherever there is a syntactical entity, there must be an observer associated with it.

This pivotal assumption, however, is exactly what is at issue. If the homunculus argument is to be taken as supporting

---

[77] The Rediscovery of the Mind, p.210
[78] The Rediscovery of the Mind, p.210
[79] The Rediscovery of the Mind, p.214
[80] The Rediscovery of the Mind, p.212

the more general claim that syntax is not intrinsic to physics, it quickly becomes clear that it cannot do this without begging the question and assuming that an observer is required if syntax is to exist. Ultimately, both claims rest on the same underlying assumption, that syntax is observer-relative and therefore cannot exist independent of observers. As a result, neither can be true in virtue of the other. The homunculus argument does not prove anything about the nature of syntax without first begging the question, and therefore only presents a different aspect of the same assumption, namely that syntax is not intrinsic to physics.

In the end, the homunculus argument only disappoints as a possible source of support for Searle's general claim that syntax is not intrinsic to physics. Perhaps Searle's analogy with functions will fare better.

## CHAPTER FOUR: THE FUNCTIONS ARGUMENT

Although the universal realization and the homunculus arguments have been shown to be lacking, Searle has one more line of support for his central premise that syntax is not intrinsic to physics. The functions argument, as I will call it, is an argument from analogy. It states that syntax is not intrinsic to physics in the same way and for the same reasons that functions are not intrinsic to physics. Certainly, Searle does not openly use the functions argument when he is attacking the computational theory of the mind. The argument's purpose is not to establish anything about syntax but instead to drive teleology out of biology. Nevertheless, it is quite clear that the functions argument can be used to support the premise that syntax is not intrinsic to physics, and, to be as charitable as possible, I will examine it as such. If functions can be shown to be extrinsic to physics, then this can only help the cause of showing that syntax is also extrinsic to physics.

### Searle's View of Functions

For Searle, the functional level of explanation is vastly unlike the hardware or the intentional levels of explanation in cognitive science. He explains how the functional level of explanation is only metaphorical because there is no separate functional level different from the causal levels. In his words, "the so called 'functional level' is not a separate level at all, but simply one of the causal levels *described in terms*

*of our interests.*"[81] In other words, functions only exist when the brute physical world is viewed with a certain bias. What is essential to a function is not only that certain physical conditions be met but also that the observer in question have certain interests with regard to the physical system. Function is observer-relative because it depends for its existence on what an observer values in the things he observes. Without an observer to impose his values or interests onto a given physical system, functions cannot exist. As Searle puts it, "where functional explanations are concerned, the only *facts* are brute, blind physical facts and the only norms are in us and exist only from our point of view."[82]

The point is not that we cannot discover functions in nature because they must be imposed by us on nature. The point is that we can only discover natural functions "within a set of prior *assignments* of value (including purposes, teleology, and other functions)."[83] Only once we as observers decide what physical features of the world we are particularly interested in can we discover functions in nature, because our discovery is ultimately dependent on the imposition of our values onto the world. As observers, we define teleological goals according to our interests, and then proceed to investigate which physical systems causally realize those goals. Functions are simply "causal relations to which we attach some *normative*

---

[81] The Rediscovery of the Mind, p.237
[82] The Rediscovery of the Mind, p.237-238
[83] The Construction of Social Reality, p.15

importance."[84] Searle uses the example of a heart pumping blood to clarify his position:

> After all, who could deny ... that the heart *functions* to pump blood? But remember when we say that the heart functions to pump blood, the only facts in question are that the heart does, in fact, pump blood; that fact is important to us, and is causally related to a whole lot of other facts that also are important to us, such as the fact that the pumping of blood is necessary to staying alive. If the only thing that interested us about the heart was that it made a thumping noise or that it exerted gravitational attraction on the moon, we would have a completely different conception of its "functioning" and correspondingly of, for example, heart disease. To put the point bluntly, in addition to its various causal relations, the heart does not have any functions.[85]

When we discover that the heart functions to pump blood, we are imposing our personal value of staying alive onto the physical world and therefore are only discovering what physical systems satisfy that value. In the end, function is not naturalistic but observer-relative, and functions only exist if an observer is there to make them exist.

The reason, according to Searle, why we tend to believe that there are natural biological functions is because we tend to take it for granted in biology that survival is valuable. Staying alive is so fundamental to us as human beings that we tend to forget that it is still a human value that we are imposing on the physical world. As Searle explains:

> If we thought the most important value in the world was to glorify God by making thumping noises, then the function of the heart would be to make thumping noises, and the noisier heart would be the better heart. If we valued death and extinction above all, then we would say that a function of cancer is to

[84] The Rediscovery of the Mind, p.238
[85] The Rediscovery of the Mind, p.238

> speed death. The function of aging would be to hasten death, and the function of natural selection would be extinction.[86]

It is only because it is so natural for us to value survival that we so easily attribute functions to the biological world. In Searle's view, any belief in the existence of natural functions is the result of imposing our interests as observers onto the interest-blind physical universe.

It is important to note that, for Searle, function, like syntax, is ontologically real even though it is observer-relative. Even though the normative component of functions rests in the eye of the beholder, once an observer recognizes a function as such, it becomes as real as brute physical features. Although our values are not intrinsic to the natural world, our values are intrinsic to us, and, once we impose functions on the world based on our intrinsic values, functions become fully real. Thus, although functions are not intrinsic to physics, they become just as real as ordinary physical features once they are imposed by observers onto the physical world. The important point is that, without observers, functions cannot exist.

## The Naturalistic View of Function

According to such writers as Larry Wright, Ruth Millikan, and Karen Neander, Searle is wrong about the observer-relativity of functions. In their opinion, functions are completely independent of observers and have nothing to do with the values or interests of human beings. Although there are slight variations between the definitions of natural function given by

---

[86] The Construction of Social Reality, p.15-16

these and other writers, the basic idea, as summarized by Wright,[87] is as follows:

> The function of X is Z means
> (a) X is there because it does Z.
> (b) Z is a consequence (or result) of X's being there.

The function of a trait, according to this view, is the effect for which that trait was selected, either by natural or agentive selection. In order to discover what functions exist, it is necessary in these philosophers' view to examine the selection-history of the world, and to isolate what effects of what traits are responsible for the current existence of those traits. Because this theory of functions focuses on past selection, and states that "what counts as a function of a trait is determined by that trait's history",[88] it is an etiological theory. According to this naturalistic definition of functions, the function of the heart is to pump blood because the effect of pumping blood forms, at least in part, the explanation for why hearts were selected for in evolutionary history. Similarly, the function of the sweep-second hand on a watch is to make seconds easier to read because the effect of making seconds easier to read explains why the hand was selected for by a conscious agent. According to these philosophers, the function of a thing is the effect of that thing that at least in part explains the thing's existence. As a result, functions clearly are independent of observers, and are instead intrinsic to and discoverable within the physical world.

---

[87] Wright, (1973), p.161
[88] Karen Neander, (1991b), p.459

Although the different theories of natural function all hold that functions are naturalistic, they do vary. For instance, whereas Wright's definition of functions is quite simple, Millikan's is much more complex. In her opinion:

> for an item A to have a function F as a "proper function," it is necessary (and close to sufficient) that one of these two conditions should hold. (1) A originated as a "reproduction" (to give one example, as a copy, or a copy of a copy) of some prior item or items that, *due* in part to possession of the properties reproduced, have actually performed F in the past, and A exists because (causally historically because) of this or these performances. (2) A originated as the product of some prior device that, given its circumstances, had performance of F as a proper function and that, under those circumstances normally causes F to be performed by *means* of producing an item like A. Items that fall under condition (2) have "derived proper functions," functions derived from the functions of the devices that produce them.[89]

Millikan's analysis of function is clearly more complicated than Wright's, showing just how different naturalistic accounts of function can be. Also, whereas Wright believes that the theory of functions can and should be expressed in terms of selection simpliciter, Neander has argued that we should specifically recognize natural selection in an analysis of biological functions and agentive selection in an analysis of agentive functions. Wright, motivated by the desire to have a univocal account of biological and artifactual functions and by the allure of remaining neutral between evolutionary theory and Creationism, leaves open what kind of selection process might apply to the ontology of a given function. Wright tries to provide a unified theory of natural functions, where functions

---

[89] Millikan, (1989), p.288

are defined in the same way regardless of the kind of selection that brings them about. Neander, on the other hand, sees this approach as overly simplistic, and offer selection-specific accounts instead. Because the selection process is more restricted where biology is concerned, she argues that the proper functions of evolved traits are more constrained than those of artifacts. For instance,

> Biological proper functions belong primarily to types and only secondarily to tokens because natural selection does not operate on individuals or their biological parts and processes. A particular piece of genetic material, or a particular instance of a trait (your thumb, Reagan's nose) cannot be selected by natural selection which operates over whole populations.[90]

Also, she points out that biological functions of evolved traits exist only in virtue of past performances of the effects that form the functions in question. The functions of artifacts, in contrast, are completely different. Because artifacts are created by agents, they first can have idiosyncratic functions that are completely unique to one artifact, and they second can have functions that never get realized, either in the past, present, or future. Unlike Wright, Neander believes that a theory of natural functions cannot be overly simplistic, but instead must accommodate the strong differences between biological and artifactual functions. Clearly, there are strong disagreements in the naturalist camp about the precise nature of natural functions, for although the proponents of the etiological theory agree that functions are natural they strongly disagree about how natural functions should be defined.

Despite these variations, the etiological theory of natural functions is an obstacle to Searle's claim that functions are observer-relative and therefore not intrinsic to the physical world. Even though there is disagreement in the etiological camp, the very possibility that functions can be grounded in the natural world poses a serious threat to Searle's position. As Searle himself acknowledges, "if such an analysis were correct, it would eliminate the observer relativity of function."[91] To combat it, Searle offers four reasons why the etiological theory is misguided. It is to these reasons that we now turn.

**Searle's Attack on the Etiological Theory**

John Searle, in The Construction of Social Reality, provides a very clear set of arguments to show that any etiological theory of functions is necessarily flawed. To begin with, he points out that such a theory "makes the definition of function dependent on a particular causal historical theory about 'reproduction.'"[92] He claims that, because philosophers such as Millikan define function in terms of evolutionary theory, they inevitably force the former to rely on the truth of the latter. Analytically speaking, the concept of function is necessarily linked to the concept of evolution, leaving us with the apparent disjunct that either evolution theory is true or there are no natural functions. According to Searle, this clearly is not the case, since functions would exist even if evolutionary theory turned out to be false. Second, Searle argues that any etiological theory of functions would seem to

---

[90] Karen Neander, (1991a), p.174

lead us to obvious counterexamples. He offers the example of the common cold, which, according to the said theory, would have to have the function of spreading cold germs. Colds, Searle insists, obviously do not have a function, and most certainly do not have the function of spreading cold germs. He claims that such counter-intuitive examples reveal just how mislead the etiological theory is. Third, Searle argues that etiological theories of function leave the normative component of functions unexplained. Even though philosophers like Millikan might be able to account for how a naturalistic function might never in fact be carried out, Searle claims that they certainly cannot account for how such functions could ever be described as better or worse, malfunctioning or functioning properly. Because the normative aspect of functions is left unaccounted for, the etiological theory must be misguided. Last, Searle argues that functions must be observer relative because naturalistic functions cannot be intensional in the way that functions obviously are intensional. In his words:

> Substitution of coreferential terms in function contexts fails to guarantee preservation of truth value. Thus "The function of A is to X" together with "X-ing is identical with Y-ing" do not imply "The function of A is to Y." For example, it is trivially true that the function of oars is to row with, and rowing consists in exerting pressure on water relative to a fixed fulcrum; but it is not the case that the function of oars is to exert pressure on water relative to a fixed fulcrum.[93]

The fact that the etiological definition of function cannot accommodate the obviously intensional nature of functions is

---

[91] The Construction of Social Reality, p.16
[92] The Construction of Social Reality, p.17
[93] The Construction of Social Reality, p.18-19

evidence that the etiological theory is misguided. Searle believes his four arguments against the etiological theory of functions are fairly convincing, and decides that the etiological theory of functions does not threaten his claim that functions are observer dependent. I will now examine the case he makes, and ascertain whether or not we have good reason to agree with him.

## The Strength of Searle's Attack

Overall, Searle's attack on the etiologists is much less successful than Searle wants us to believe. Upon close inspection, it becomes clear that three of the four arguments he provides carry very little argumentative weight. His claim that the etiological theory leads us to obvious counter-examples of function is especially laboured. The one example that Searle provides, that the function of colds would on the etiological theory be to spread cold germs, is not at all an obvious counter-example. On the contrary, looking back at Wright's definition of function, it seems perfectly plausible to say that colds exist because they spread cold germs and that cold germs are spread as a result of colds existing. Searle counts on his example to seem obviously ludicrous, but I see no reason why it should be. Whether or not colds have that function may still be difficult to determine, and an answer to that question may very well require some honing of the meaning of the ambiguous terms involved. Even so, that does not mean that natural functions are shown to be preposterous. The only way that this argument carries any weight is if the idea that colds might have the

function of spreading cold germs is absurd, and I see no reason to believe that it is. Therefore, I will drop this argument from consideration.

Searle's charge that natural functions leave the normative component of functions unexplained also seems to come up short. According to Searle, the etiological theory of functions cannot account for how we can speak of better or worse hearts, or of hearts that malfunction. I see no reason why we should believe this. As Neander explains, hearts clearly function to pump blood because "they are all *supposed* to pump blood; by which I mean that pumping blood is what they were selected for--it is their proper function."[94] Hearts can sensibly be said to malfunction because the biological notion of function can still be a teleological notion, even though it is routed in evolutionary history. According to Neander,

> talk of functions involves forward-reference to the effects that items or traits are supposed to have, and also an implicit backward-reference to a causally explanatory selection process, during which those items or traits were selected for those effects which are their functions.[95]

This analysis has the result that traits can be unproblematically dysfunctional. In the end, "dysfunctinal traits are dysfunctional precisely because they have functions that they are supposed to perform, but which they lack the disposition to perform."[96] Under this interpretation of functions, it is unproblematic and certainly accounted for how hearts can be good, bad, properly functioning or malfunctioning

---

[94] Neander (1991b), p.467
[95] Neander (1991b), p.467
[96] Neander (1991b), p.466

hearts. The normative element of functions is explained by the fact that functions can be teleological as well as etiological. A functional description of a heart can therefore be a description of how hearts function normally. Searle assumes that norms can only be provided by intentional observers, but I see no reason why they cannot be provided by evolution itself, and so I will let this argument drop as well.

Searle's argument that etiological functions cannot be intensional encounters the same fate. Searle claims that the function of oars is clearly to row and clearly not to exert pressure on water relative to a fixed fulcrum, even though that is precisely what rowing is. This argument only has bite, however, if we already concede that functions are attributed by observers. We only have reason to believe that functions are intensional if we already believe that they are imposed onto the world by intentional agents. What the etiologists are claiming, however, is that functions are real features of the world that can be discovered and identified like any other real feature. According to Millikan, functions are theoretical entities, and, as such, exist regardless of how we conceptually identify or perceive them. If the function of A is to X, this is true only because A was selected for as the result of its capacity to X. If it turns out that X's are identical to Y's, then the function of A would also be to Y for the exact same theoretical reasons that originally made X the function in question. Because it is the underlying reality and not the way we view that reality that is important, functions, according to the etiologists, are not

intensional at all. This argument only supports Searle's theory of functions if that theory is already presupposed, and, because of this, it also fails as an argument against the etiological theory of functions.[97]

The one strong argument that Searle poses against the etiological theory is that such a theory makes the definition of function dependent on the truth of evolutionary theory. Obviously, it is undesirable to tie the definition of a thing to the truth of a scientific theory about evolutionary reproduction. As Searle puts it, "even if no such account of reproduction, Darwinian or otherwise, turned out to be true, my heart would still function to pump blood."[98] Analyzing the notion of function in terms of natural selection might even be said to be analytically arrogant because it begs empirical and theological questions.[99] Such an analysis would entail from the obvious fact that functions exist the undesirable conclusion that the theory of natural selection is necessarily rather than contingently true, which of course it is not. If function presupposes natural selection, then if functions exist we are forced to assert that natural selection must exist also. Certainly, it would seem to be a grave problem if natural functions imply, rest upon, or logically entail the truth of evolutionary theory.

---

[97] Bruce Hunter has pointed out that this third argument of Searle's can also be undermined by noting that Searle's assumption that causal contexts are extensional is contentious.

[98] The Construction of Social Reality, p.17

[99] Neander (1991a), p.176-177

This argument is supported by a point raised by Earnest Nagel, Christopher Boorse, and Larry Wright.[100] They claim that the etiological theory of function is flawed because defining functions in terms of evolutionary theory is ahistorical. As they point out, the discovery of biological functions historically predates the discovery of natural selection. When Harvey declared that the function of the heart was to circulate blood in 1616, he clearly did not have natural selection in mind. One has to wonder how, if proper functions are really definable in terms of evolutionary theory, the theory of functions can predate Darwinian theory by two centuries. As Neander explains,

> Presumably, logically derivative notions must be chronologically antecedent notions. So, the etiological theory is charged with the implication that natural selection was discovered before the notion of a (biological) "proper function" was used, and this is plainly false.[101]

Whereas Searle's point is that defining function in terms of natural selection is precarious and even counter-intuitive because it makes functions dependent on the truth of evolutionary theory, the point made by these philosophers is that it is also historically inaccurate. Together, the two arguments stand as a formidable obstacle for any etiological theory of functions.

How might the etiologist escape this attack? One possibility would be to simply bite the bullet and accept the charges. Although it would put the etiological theory in a

---

[100] Wright (1976), p.96-97, Boorse, p.74, Earnest Nagel, p.284
[101] Neander (1991a), p.175

theoretically awkward and historically backward position, the etiologist could simply assert that a naturalistic account of functions has to presume the truth of evolutionary theory, and he could make the claim that Harvey in fact did not truly discover the function of the heart because he was unaware of the necessary background theory. It would seem that Millikan, in her belief that the etiological theory is a theoretical theory subject to the truth of its theoretical framework, would give this response. As ahistorical as it might appear, the true definition of function is essentially tied to the theory of natural selection, and, regardless of what Harvey thought, he did not discover that the function of the heart was to pump blood because he did not know that, as it turns out, functions are the effects for which traits were selected for by natural selection.

Neander, however, launches a much more aggressive defence of the etiological theory. She interprets the etiological theory as a conceptual rather than as a theoretical theory, and believes that the etiological theory of function is tied to the truth of evolutionary theory in a time-transcending, unproblematic way. Because she believes that the background theory of evolution might very well change, she acknowledges that the biologist's understanding of the notion of natural function might very well shift dramatically. She is very open to the possibility of scientific revolution, and she accepts the fact that scientific terms are shaped by background theories and that these background theories are subject to change. She

argues, though, that this is not problematic for the etiological theory because the intention of the theory is to ground the definition of function not in any current biological theory but instead in an ever-changing, timeless biological theory. Neander does not believe that the etiological theory entails that either natural selection theory is true or there are no proper functions, but instead that "natural selection theory is true or there are no proper functions *in the modern biologist's sense of 'proper functions'*."[102] Even if Darwinian evolution turns out to be false and is replaced by another theory, Neander explains that the concept of function is flexible enough to adapt to the new theory. As she puts it, "even if the theory of natural selection is falsified, and the etiological theory *as theoretical definition* is thereby shown to be untrue, some other theoretical definition of proper functions will be true."[103] In support of her position, Neander gives the example of time as it was once understood under the background theory of Newtonian physics. She explains:

> Now we can ask whether this would entail that either time is absolute or there is no time. And it is obvious that it does not. Rather it entails that either time is absolute or there is no time in the Newtonian physicist's sense of "time".[104]

Neander has faith that functions will undoubtedly remain grounded in the natural world no matter how the background theory of evolution changes. As she puts it,

> Newtonian physics was indeed falsified, and so the Newtonian theoretical definition of 'time' was falsified. But some other theoretical definition of

[102] Neander (1991a), p.177
[103] Neander (1991a), p.177
[104] Neander (1991a), p.177

> 'time', or at least of 'space-time' is presumably true. (Time still exists.)[105]

When Harvey discovered the function of the heart, the selection process that he had in mind, presumably the design of God, was different from the selection process currently favoured, namely, natural selection. According to Neander, the reason why this is unproblematic is because Harvey supposed that biological functions were the result of *some kind of selection process*. The fact that he believed that the selection process was the design of God is non-essential. Therefore, even though scientific revolutions have taken place and the selection process commonly accepted has changed, Harvey nevertheless captured the spirit or the essence of natural functions, and that is enough. In the end, according to Neander, Searle's argument that the definition of function is awkwardly tied to the truth of evolutionary theory and is therefore counter-intuitive and the supporting argument that such an analysis is ahistorical and therefore dubious both fail to hurt the etiological theory because the etiological theory is not theory-specific but instead theoretically dynamic.

What are we to make of this dispute? In the end, it would seem that even Searle's strongest argument against the etiological theory of functions is less than successful, and so his attack against the possibility that there might be naturalistic functions ultimately falters. In the first case, if the etiologist follows Millikan and simply asserts that the definition of function is based on a background theory of

evolution, it would seem that we ultimately come to a cross-roads of intuitions. Certainly there is a certain amount of risk and perhaps even awkwardness in requiring the truth of evolutionary theory in order to make functions naturalistic, but it could be argued that there is also some risk and awkwardness in Searle's own theory of functions. Whereas Millikan's definition of function ultimately rests on the truth of the theory of natural selection, Searle's definition of function ultimately rests on the truth of his theory about intrinsic and observer-relative features in the world. One could argue that Searle's theory problematically implies the disjunct that either his intrinsic/observer-relative theory is true or there are no functions. To mirror Searle's own argument, we could respond to this that even if Searle's theory turned out to be false, there would nevertheless be functions. Also, Searle's own theory would seem to be as historically awkward as the etiological theory. Harvey discovered the function of the heart nearly *four* centuries before Searle made public his theory that there are intrinsic and observer-relative features in the world. Although Harvey did not have the background theory of natural selection in mind when he said that the function of the heart is to circulate blood, he also did not have Searle's theory in mind. It is not at all clear that Millikan's stance on functions depending on the truth of evolutionary theory is especially problematic, and so Searle's strongest argument against etiological theory is at worst a minor nuisance.

---

[105] Neander (1991a), p.178

Furthermore, if Neander's more grand interpretation of functions is upheld, Searle's attack on the etiological theory all but peters out. Although assessing Neander's position seems especially difficult and deserves further enquiry, she does provide a neat rebuttal to Searle's best argument against the etiological theory of functions. It would seem that, where Searle's attack on naturalistic functions is concerned, the attack is shown to be either limited or entirely lacking. In order to show that functions cannot be naturalistic but instead must be observer-relative, and therefore to show that we ultimately have good analogical evidence for the claim that syntax too is not naturalistic but instead observer-relative, Searle must show that the etiological theory of functions is inferior to his own theory. In fact, he does not succeed in doing this. In the end, we are left with the impression that the etiological theory is at the very least plausible, and this is enough to weaken Searle's analogical evidence. The etiological theory of functions certainly may be misguided, but there is as much, if not more, reason to think that it is on the right track than to think that it is not.

Thus, the functions argument, taken as an argument from analogy for the claim that syntax is not intrinsic to physics, is shown to be lacking. When all is said and done, it seems at least as reasonable to say that functions are intrinsic to the natural world as it is to say that functions are observer-relative and therefore nonnatural. Along with the universal realization argument and the homunculus argument, the functions

argument fails to provide any convincing reason why we should accept Searle's central premise in his attack on the computational theory of the mind. Whether or not syntax is actually intrinsic to physics is therefore left unsupported. In the following chapter, I will examine what consequences follow from this.

## CHAPTER FIVE: THE NATURE OF SYNTAX

As has been shown in the past three chapters, Searle fails to provide a convincing argument for his central premise that syntax is not intrinsic to physics. The universal realization argument, homunculus argument, and functions argument all fail to establish the key claim that syntax is observer relative and therefore extrinsic to the physical world, and, because Searle's attack on the computational theory of the mind depends on the truth of that premise, we are left without any compelling reason to reject cognitivism, the claim that the brain is a digital computer. Although Searle believes that syntax is obviously dependent on observers for its existence and therefore that syntactically defined entities cannot be observer-independent, he cannot provide us with a convincing reason why we should share that belief. Because of this, his overall claim that the computational theory of the mind is fundamentally flawed is left wholly unsupported.

Not only is the claim left unestablished, but it also has its own potential difficulties, making Searle's attack on cognitivism not just unsupported but also potentially problematic. As will be shown, Searle's central premise is not just ill-defended but also laden with potential problems, giving us not only no concrete reason to accept it, but also possible reason to flat out reject it.

## Intrinsic to Physics

Many of the criticisms brought against Searle's attack on cognitivism centre on Searle's use of the phrase "intrinsic to physics". One criticism with the use of this phrase as a criterion for a thing's observer-independent existence is that it assumes an intolerably reductionist ontology. According to Roger Fellows, when Searle uses 'intrinsic to physics' as a criterion for existence, he implies that a thing needs to be reducible to physics in order to exist. Fellows believes that, if a philosopher is not willing to become a strong reductionist, he is going to have a problem with Searle's requirement that something be intrinsic to physics in order to exist independently of observers. He gives the example of a person discovering a pattern of stones on a hillside that spell out *Welcome to Wales*:

> If the stones rolled down the hill and formed the pattern according to the erosion of the rocks and so on, then the explanation would lie in physics. However, one can imagine (albeit fancifully) a biological explanation in terms of a group of badgers having moved the stones around for the purposes of their own. The explanation for the resulting pattern is extrinsic to physics and lies within the biological domain.[106]

Fellows believes that Searle's attack on cognitivism hastily excludes the possibility that syntax might be entirely irreducible to physics and yet completely naturalistic. He continues:

> In general, unless one espouses the position of the hardline reductionist who insists that all genuine explanations must be couched in the language of physics, we have to acknowledge that there are many

[106] Fellows, p.96

> explanatory levels: the physical, the chemical, the biological, the psychological, and so on, no one of which can be reduced to the others in any interesting sense.[107]

Fellows' charge, then, is that the criterion of being intrinsic to physics is simply too strong. According to Fellows, it is perfectly reasonable to expect that many completely naturalistic phenomena might fail such a severe condition. Searle's attack on cognitivism thus incorporates a fairly strong reductionist ontology, limiting its appeal to those who share that ontology.

Another problem that has been identified in requiring that syntax be intrinsic to physics is that the requirement wrongly presumes that physics, as a science, can be suitably used as a measuring stick of reality. According to Ranjan Umapathy, physics is not nearly the closed and unified science that it would need to be in order to fill such a role. In his opinion, the physics community is constantly in a state of flux, making its successes and failures extremely difficult to gauge. As he explains,

> The state of theoretical physics today seems entirely indecisive. Nobody seems to know what the physical world is really like--physicists do not seem to have reached a consensus on what seems to be the simplest question, 'what is an atom, really?', let alone about 'quarks' and 'mesons'. I am inclined to think that all the technological advances that have been attained are not because we have a great and decisive theoretical grasp of the physical universe, but because we have been able to, thanks to experimentation, discover certain mathematically describable relations that occur in the physical universe.[108]

---

[107] Fellows, p.96
[108] Umapathy, p.45

Umapathy claims that Searle's use of physics as a measure of reality presupposes a high degree of unity and accomplishment in the field of physics, and, in Umapathy's opinion, the world-view of physics is not sufficiently unified or established to be included in a criterion of existence. Thus, another problem that has been attributed to Searle's claim that syntax is not intrinsic to physics is that is presumes that physics is a science that is unified and well-founded enough to warrant its use as a reference point in gauging the reality of things like syntax.

I do not think that the two previous points are very problematic for Searle. The use of the word 'physics' certainly is not crucial to Searle's overall program; he could easily talk about being intrinsic to the natural world instead. Criticisms based on the inadequacy of physics or on the decision to talk in terms of physics at all could easily be deflected by changing "syntax is not intrinsic to physics" to "syntax is not intrinsic to the natural world". At best, I believe that Fellows and Umapathy point out an ambiguity in Searle's project, and not a serious problem with it.

There is, however, a more serious difficulty with Searle's requirement that a thing be intrinsic to physics in order to exist. The difficulty is that Searle's requirement potentially conflicts with his own theory of the irreducibility of consciousness. As Beth Preston[109] points out, Searle argues that consciousness should be understood as a system feature of the

[109] Preston, p.142-143

physical brain, one which causally emerges from the system of neurons therein. According to Searle, the subjective level of description of the brain cannot be reduced to the objective level because consciousness consists in its appearance, and, where the appearance of consciousness is concerned, "*we cannot make the appearance-reality distinction because the appearance is the reality.*"[110] In his work on consciousness, Searle puts forward and defends the view that the first-person subjective level of description, or the appearance, of consciousness, need not be reduced to the physics of the natural world in order to be viable. He defends the plausibility of there being a subjective level of reality that is not definable in purely objective physical terms, and he even claims that this irreducibility "has no deep consequences".[111]

Where syntax is concerned, however, Searle seems to put forward the exact opposite view. Whereas consciousness can be an intrinsic property of a physical system despite the fact that it is not a property at the level of description of physics, syntax apparently cannot be an intrinsic property of a physical system specifically because it cannot be physically defined. As Preston puts it,

> Searle's position is that brains *are* minds, *in spite* of the fact that the subjective and objective levels of description are not interdefinable; but brains *are not* computers, precisely *because* the syntactic and physical levels of description are not interdefinable.[112]

---

[110] The Rediscovery of the Mind, p.122
[111] The Rediscovery of the Mind, p.118
[112] Preston, p.142-143

In his attack on cognitivism, Searle clearly assumes that a property of a system, if it is to be an intrinsic property, must be a property of it at the level of description of physics. Searle specifically claims that syntactical entities cannot be observer-independent because where syntax is concerned "the physics is irrelevant"[113]. In his philosophy of mind, however, Searle defends the existence of consciousness even though it too cannot be described at the level of description of physics. According to Searle, even though consciousness is brought about by physical events and is rooted in a causal micro-structure, we cannot reduce consciousness to physics because where consciousness is concerned we are only interested in its epistemic bases, i.e., its appearance. Searle opens the door to the possibility that a thing's ontology can be defined by its appearance rather than by its physical constitution. With respect to syntax, however, this possibility is suddenly unavailable. Syntax is required to be intrinsic to physics, and any account of syntax is expected to be a physical account. When Searle claims that the irreducibility of consciousness has no deep consequences, he means that consciousness is intrinsic to the world even though its appearance is its reality. Ironically, his attack on cognitivism is based on the belief that the irreducibility of syntax does have deep consequences, since syntax is not intrinsic to the world precisely because the appearance of syntax is its reality. As Preston explains, syntax and consciousness seem required to share the same fate:

---

[113] The Rediscovery of the Mind, p.207

> Either mental states are emergent, intrinsic properties of brains, in which case syntactic states could be emergent, intrinsic properties of brains as well; or the dual-description problem is ontologically decisive, in which case neither mental states or syntactic states are real properties of brains.[114]

Thus, Preston charges Searle with inconsistency, and claims that if consciousness can consist of its appearance then so can syntax.

Again, I do not think that Searle would be terribly hamstrung by this criticism. Searle would only have to point out that, on his own theory, consciousness *could* have been ontologically reduced to its physical micro-structure. It is only because it is the appearance of consciousness that interests us that we do not do so. Syntax, Searle might point out, could *never* have been reduced to any physical properties, and it is for that reason that his position is consistent. Put simply, consciousness has a physical make-up, and syntax does not. Just because we decline to define consciousness in terms of that makeup does not mean that it does not have one. Searle's rebuttal would probably be that consciousness is ontologically real because it is grounded in the natural world, whereas syntax is not because it is grounded only in the eyes of observers.

I believe, however, that Preston's criticism is not entirely off-base. In my opinion, the important point is that Searle has set a precedent where epistemic bases can be used to define naturalistic phenomena. Searle claims that the irreducibility of consciousness has no deep consequences, but

that is false. The deep consequence is that it is suddenly possible to define a thing in terms of how it subjectively appears instead of in terms of how it objectively is. Searle claims that syntax is not intrinsic to the physical world because it is observer-relative and only exists in the eye of the beholder. In a very real sense, however, consciousness is also observer-relative, or at least agent-relative, for it too only exists in the eye of a beholder. Although consciousness has a micro-structure, the micro-structure alone is insufficient for the existence consciousness. What we need in order for consciousness to exist is the subjective experience of consciousness by a conscious agent. One might argue therefore that syntax and consciousness are in the same boat; both of them require an observer to experience them in some fashion in order for them to exist. Searle says that syntax is not naturalistic because the existence of syntax is essentially tied to an observer's attribution of it. Without fear of contradiction, he also says that consciousness is naturalistic, even though the existence consciousness is also essentially tied to an observer's experience of it. Once epistemic bases are tolerated as bases for existence, it seems difficult to assert that syntax is not intrinsic to the natural world because it is fundamentally linked to its epistemic bases. Searle argues that if there had never been any observers syntax would never have existed because syntax needs to be attributed by an observer. What he overlooks, it would seem, is the fact that if conscious

---

[114] Preston, p.143

agents had never existed, consciousness would never have existed because consciousness needs to be experienced by an agent in order to exist. Thus, the criticism remains. When Searle claims as a central premise that syntax is not intrinsic to physics, Searle's own theory of consciousness implores one to ask "and why does it need to be?"

I will not pursue this line of argument any further, not because I do not think it can be pursued, but because I personally think it is the wrong direction to take on the issue. In response to Searle's claim that syntax is not intrinsic to physics, Fellows and Umapathy ask why something should need to be intrinsic to *physics* in the first place, and Preston asks why syntax but not consciousness should be required to reduce to the physical world. In other words, they all downplay the *need* for syntax to reduce to physics, and ignore the question of whether or not it actually *does*. I intend to break from that trend. My strategy will be not to reject but to accept the criterion that something needs to be intrinsic to physics, or more generally to the natural world, in order to ontologically exist, and to instead explore what problematic consequences follow when we accept the claim that syntax is not intrinsic to physics.

## Slippery Slope

A concrete problem with Searle's claim that syntax is not intrinsic to physics is that it threatens the observer-independent existence of other components of the external world. Because the premise would seem to imply not only that syntax but also much of the external world is observer dependent, accepting

syntax as observer-relative is problematic because doing so requires us to accept that other aspects of the external world must be observer-relative as well. The line that Searle draws between intrinsic and observer-relative seems precariously placed, and a slippery slope appears to threaten. Searle's central premise is extremely potent. It would seem that if we accept Searle's claim that syntax is extrinsic to physics then for the very same reasons we accept it we would have to accept that many other things in the external world would also have to be extrinsic to physics.

Roger Fellows provides a good example of how Searle's premise might spill over to other aspect of the physical world. In his words:

> consider the fact that chemical theory *conventionally* assigns the symbols, 'H' and 'O' to a particular substance, so that for instance, water is identified as having a certain structure. Nevertheless it would remain true that water would equal $H_2O$ even if Humanity disappeared off the map.[115]

Certainly a parallel can be drawn between our assignment of computational syntax to physical phenomena and our assignment of a labelling system to various chemicals. Fellows argues that chemicals, even though we assign a labelling syntax to them, are completely intrinsic to the world. Although the actual term 'hydrogen' or 'oxygen' would not have existed if there had never been any observers, this has no bearing on whether hydrogen and oxygen would have existed. Because conventionally assigned labels have nothing to do with the ontology of the reality that

---

[115] Fellows, p.96

is labelled, we therefore have reason to think that neither chemicals nor syntax are observer-relative phenomena.

Searle, of course, is a strong external realist, and would agree with Fellows that chemicals exist externally from our representations of them. According to Searle, something is intrinsic to the world if it would exist even "if there had never been any human beings or other sorts of sentient beings"[116] to observe it. Although Searle concedes that it may at times be difficult to determine whether something is intrinsic to the world or not, as in the troublesome case of colors,[117] molecules certainly qualify, and syntax certainly does not. He explains:

> It is, for example, an intrinsic feature of the object in front of me that it has a certain mass and a certain chemical composition. It is made partly of wood, the cells of which are composed of cellulose fibers, and also partly of metal, which is itself composed of metal alloy molecules. All these features are intrinsic.[118]

In Searle's unabashed external realism, the existence of chemicals like oxygen and hydrogen are taken not only to be intrinsic to the world, but as obviously intrinsic to the world. Searle's confidence in this fact is extreme:

> Always ask yourself what we know about how the world works in fact. It might have turned out differently but this is in fact how it did turn out: the universe consists entirely of particles in fields of force. These particles are organized into systems. Some of these are natural systems, such as galaxies, mountains, molecules, and babies. ... Among the natural systems, some are living organic systems. They contain carbon-based molecules and have heavy doses of nitrogen, oxygen, and hydrogen.[119]

---

[116] The Construction of Social Reality, p.11
[117] The Construction of Social Reality, p.11
[118] The Construction of Social Reality, p.9
[119] The Mystery of Consciousness, p.213

Unlike syntax, which in his opinion depends entirely on the prior existence of observers in order to exist, the objects of chemistry are for Searle clear instances of intrinsic physical objects. They epitomize observer-independence, and stand in stark contrast to observer-relative entities such as computational syntax.

When Searle's own reasoning is turned against him, however, it is not at all clear that molecules are the intrinsic objects he believes them to be. Recall Searle's claim that we make the mistake that biological functions are intrinsic to the world because we assume that survival is inherently valuable, and impose this value onto the biological world. We attribute functions in biology only once we impose our own interests as observers onto the interest-blind physical universe. In the case of the science of chemistry, it might be argued that it was only after scientists imposed their values of regularity and pattern formation that they were able to form the periodic table of elements. Certainly, scientists formed the periodic table because the table has explanatory power, and because it was suitable for the prediction and categorization of experimental phenomena. Nonetheless, just as Searle argues that it is only after we impose our interest in survival on biology that we can discover biological functions, one might similarly argue that it is only after we impose our interest in experimental regularity and patterns between those regularities that we can discover molecules such as hydrogen and oxygen. The elements of chemistry, it might be argued, have as much to do with the

interests of human beings as natural functions in biology and even the syntax of computers. Thus, the distinction between what is intrinsic and what is observer-dependent becomes blurred, because, for the same reason that functions are observer-dependent, entities like oxygen and hydrogen would seem to be observer-dependent as well.

In response to this, Searle might argue that, no matter what we might say about our science or classification scheme of molecules, the underlying reality, whatever that might be, exists nevertheless. Independent of what observers believe or ascribe, molecules *of some sort* are intrinsic to the physical world. Searle does not think that arguments from conceptual relativity have any bite against external realism. Even though we necessarily incorporate an arbitrary set of concepts and an arbitrary vocabulary when we describe a thing, the thing in itself is independent of our concepts or vocabulary. He gives the example of the Himalayan mountains:

> Take a corner of the world, say, the Himalayas, and think of it as it was prior to the existence of any human beings. Now imagine that humans come along and represent the facts in various different ways. They have different vocabularies, different systems for making maps, different ways of counting one mountain, two mountains, the same mountain, etc. Next, imagine that eventually the humans all cease to exist. Now what happens to the existence of the Himalayas and all the facts about the Himalayas in the course of these vicissitudes? Absolutely nothing. Different descriptions of facts, objects, etc., came and went, but the facts, objects, etc., remained unaffected. (Does anyone really doubt this?)
>
> The fact that alternative conceptual schemes allow for different descriptions of the same reality, and that there are no descriptions of reality outside all

> conceptual schemes, has no bearing whatever on the truth of realism.[120]

According to Searle, all of the possible conceptual schemes an observer can apply to the world will realize consistent descriptions of that world. Applying different conceptual schemes to the Himalayas is as unproblematic as weighing oneself in pounds or in kilograms. We might arrive at different conceptual facts with different possible schemes, but this is unproblematic. Searle, for instance, weighs either 160 pounds or 73 kilograms,[121] depending on which measuring scheme we apply to gravitational attraction of masses on the surface of the Earth, but these two facts do not show any inconsistency, nor do they threaten external realism. Searle's weight remains intrinsic to the world, and would exist even if no observers had ever been around to weigh him.

According to Searle, any argument from conceptual relativity against external realism inevitably involves the commitment of a use-mention fallacy. He explains:

> From the fact that a *description* can only be made relative to a set of linguistic categories, it does not follow that the *facts/objects/states of affairs,/etc., described* can only *exist* relative to a set of categories. Conceptual relativism, properly understood, is an account of how we fix the application of our terms: What counts as a correct application of the term "cat" or "kilogram" or "canyon" ... is up to us to decide and is to that extent arbitrary. *But once we have fixed the meaning of such terms in our vocabulary by arbitrary definitions, it is no longer a matter of any kind of relativism or arbitrariness whether representation-independent features of the world satisfy those definitions, because the features of the world that*

[120] The Construction of Social Reality, p.164-165
[121] The Construction of Social Reality, p.165. This is Searle's own example.

*satisfy or fail to satisfy the definitions exist independently of those or any other definitions.*[122]

The interests of human beings are therefore not relevant to external reality the way that they are relevant to syntax or function. How we define molecules like hydrogen and oxygen may be completely up to us and dependent on our interests as scientists and observers, but the fact that molecules of some sort exist and have intrinsic features is observer-independent. It might be up to us how we choose to describe molecules or mountains, but certainly it is an intrinsic fact about the world that, independently of how we describe them, molecules and mountains exist. According to Searle, the line between observer-dependent and observer-independent is easy to draw; syntax and function exist only relative to observers, and the external world exists independently of observers.

As confident as Searle is about the truth of external realism, and as sure as he is that there is a clear distinction between intrinsic and extrinsic, the inconsistency remains. If we look carefully at Searle's own words, we find him saying that once we fix the meaning of our terms according to our arbitrary definitional interests, there is a clear fact of the matter whether those terms coincide with the representation-independent facts of the external world. He claims that once we define 'mountain' or 'kilogram', it is completely non-relative and non-arbitrary whether or not the Himalayas exist and whether Searle weighs 73 or 75 kilograms, and so although our labelling scheme might be observer-relative, the underlying reality certainly is

not. In his discussion of functions, however, Searle specifically says that *it is because we impose our arbitrary definitional interests in the first place* that functions are observer-relative. According to the precedent set by Searle with respect to function, regardless of whether or not a neat correspondence is realized between the external world and the particular labelling scheme we choose to apply to it, the fact that labelling schemes are set by observers at all makes the things labelled observer-dependent. By Searle's own reasoning, because our interests as observers are crucially involved in our schematization of the external world, the external world must be as observer-relative as functions and syntax. When Searle tells us what we know for sure, he begins with the fact that the universe consists entirely of particles in fields of force. He believes that this claim, or some similar claim, is true regardless of whether any observer knows it. However, on careful insight, it would appear that this claim, and all claims that resemble it, all involve the same imposition of observer interests onto the world as the claim that the function of the heart is to pump blood. That the external world in itself would be identical to itself with or without observers seems uncontroversially true. That "all the facts about the Himalayas"[123] would remain true with or without observers, on the other hand, seems problematic, because such facts seem essentially tied to the interests of observers. Searle says that, although there are different ways of counting and

---

[122] The Construction of Social Reality, p.166

labelling the Himalayas, the Himalayas themselves intrinsically exist. It would seem, however, that it is only with respect to our interest in large geological formations with tall singular summits and visible silhouettes that the Himalayas exist. Recall that, for Searle, "the *discovery* of natural function can take place only within a set of prior *assignments* of value".[124] It would seem that the discovery of the Himalayas, too, can only take place within a set of prior assignments of value. Searle's assertion that once we decide on a definitional/descriptive scheme it is a matter of intrinsic fact whether the features of the world satisfy that scheme is undermined by his prior assertion that how we decide on such a scheme crucially involves our interests as observers and therefore makes the thing schematized ontologically observer-dependent. There is a strong difference between saying that the world would be physically the same as it is now even if there had never been any observers and saying that the Himalayas or oxygen molecules would exist even if there had never been any observers. Once you open the door and say that the existence of syntax or function is essentially dependent on the attitudes of observers, it becomes difficult to say that the existence of mountains or molecules is not, and although Searle tries to block that conclusion, his attempt runs into inconsistency. Therefore, because our interests as observers are essentially involved with other aspects of reality and not just with function or syntax, it would seem that many

---

[123] The Construction of Social Reality, p.164
[124] The Construction of Social Reality, p.15

more things than just function or syntax must be observer-dependent and therefore extrinsic to the physical world.

Thus, Searle's central premise that syntax is not intrinsic to physics seems to imply that much of the external world is also not intrinsic to physics, and is therefore problematic. If we accept Searle's claim that syntax is observer-dependent, we seem unable to escape a slippery slope toward the position that most if not all of the external world is also observer-dependent. This undesirable result can only count as a strike against the premise itself. Clearly, syntax and function cannot be observer-dependent the way Searle believes, because if they are then it would seem to follow that much of the external world is observer-dependent as well. Searle's line in the sand between intrinsic and extrinsic seems precariously drawn. Because accepting that syntax and function are observer-relative requires us to accept that much of the external world is observer-relative as well, we would seem to have good reason to reject Searle's central premise.

## All That Is Required

I have not exhaustively pursued these difficulties with Searle's central claim that syntax is not intrinsic to physics partly because of time constraints but chiefly because they are unessential for the success of my general program. The purpose of this thesis is to assess the strength of Searle's attack on cognitivism, on the theory that the brain is a digital computer. All that is required to show that Searle's attack on cognitivism fails is for the central premise in Searle's argument to remain

unestablished. Andrew Melnyk, in his attack against Searle's Chinese Room argument, makes a similar point:

> ... we can say that *Searle's abstract argument against strong AI is inconclusive because it rests upon a definitional assumption that we have been given no reason to accept.* Notice that this objection to Searle does not require proving the *superiority* of the alternative account, since the point is that the account Searle assumes is undefended, not that it is definitely false; so the mere plausibility of an alternative account is enough.[125]

My aim will be to provide just such an alternative account, one which avoids the difficulties posed in this chapter, and one which I feel arrives at a more intuitive position. Although I will claim that this alternative account is superior to Searle's, any superiority is merely icing on the cake. Because Searle cannot provide a good argument for his main premise, all that is required to collapse his attack is that the alternative account be at least as plausible. If it proves to be more plausible, this is only a bonus.

## My Alternative Account

I believe that the solution to the problem of the nature of syntax is actually hidden in one of Searle's own comments. In The Rediscovery of the Mind, Searle briefly mentions the possibility of talking about "the existence of systems with discrete physical phenomena".[126] He claims that it is no help to talk of such phenomena in place of talking about syntax because such phenomena are just as observer-relative as syntax and therefore just as extrinsic to physics. Even if you stop using syntactical tokens such as 1's and 0's, and talk instead about

---

[125] Andrew Melnyk, p.408

the underlying computation, the abstract algorithms or programs which syntax is used to express, you cannot escape the general problem. Searle explains:

> A physical state of a system is a computational state only relative to the assignment to that state of some computational role, function, or interpretation. The same problem arises without 0's and 1's because *notions such as computation, algorithm, and program do not name intrinsic physical features of systems.*[127]

My claim will be that Searle is dead wrong on this point. Although token syntaxes are dependent on observers to fix them onto the physical world, the underlying syntactical phenomena to which token syntaxes are attached are completely observer-independent. While the labelling scheme for syntactical phenomena is observer-relative, the phenomena themselves are not. The same is true of functions. Although token functions are closely linked to the labelling schemes of biologists, the underlying intrinsic functions exist independent of biologists, and existed before any biologist discovered them. In my opinion, underlying syntactical phenomena need to be distinguished from token syntaxes and identified as the phenomena that can be justly described and labelled with any one of a number of possible token syntaxes. A physical system is syntactical if and only if a token syntax can be consistently, accurately, and honestly applied to it. My realism is therefore more all-encompassing than Searle's; whereas he excludes syntax and function from the set of real, observer-independent things in the world, I include them.

---

[126] The Rediscovery of the Mind, p.210
[127] The Rediscovery of the Mind, p.210

In my opinion, Searle, in his attack on cognitivism, is guilty of a kind of use-mention fallacy of his own: from the fact that a description of syntactical phenomena can only be made relative to an observer, it does not follow that the phenomena themselves can only exist relative to an observer. The 1's and 0's in a desktop computer are surely observer-dependent; they would not exist if an observer did not ascribe them to the machine. Nevertheless, the syntactical binary structure that we as observers describe when we talk about 1's and 0's is observer-independent. If observers were to suddenly appear in a computer-ridden universe and apply a syntactical labelling scheme to those computers, their act, in my opinion, would be as much an act of discovery as the act of labelling the periodic table or of mapping out the Himalayan mountains. The underlying reality behind syntax is independent of the labelling scheme, and to charge that syntax is dependent on observers is to confuse the use with the mention of syntactical phenomena.

Searle claims that talk of discrete physical phenomena is of no help because notions such as computation and algorithm do not name intrinsic features of systems. Something is a computation or algorithm only in virtue of an observer's belief about it. I would argue that we have no good reason to believe this. Jack Copeland provides an observer-independent definition of algorithm that seems perfectly reasonable:

> An algorithm is a 'mechanical' or 'moronic' procedure for achieving a specified result ... . That is to say, an algorithm is a finite list of machine executable instructions such that anyone or anything that correctly follows the instructions in the

> specified order is certain to achieve the result in question.[128]

With this definition set, computation is easy to define: "to compute is to execute an algorithm."[129] Searle believes that it is of no use to talk about abstract concepts such as algorithm and computation because, in his opinion, these concepts are as observer-dependent as token 1's and 0's. Copeland, however, provides perfectly naturalistic definitions of these concepts. Is there any reason why we should side with Searle on this point?

It would seem that, in addition to his complaints about syntax being extrinsic to physics, Searle rests his case on as many as two new assumptions here. First of all, Searle clearly seems to be assuming that 'formal' implies 'observer-relative'. Louise M. Antony sees this assumption very clearly in Searle's writing, and sees no reason why it should be upheld:

> The crucial inference--from the claim that a property is *formal* to the claim that it is observer-relative--is simply invalid. The fact that a checkerboard contains alternating light and dark squares doesn't depend on my seeing it that way. Why think it's so for computational structure?[130]

Indeed, there seems to be no good reason for us to accept this assumption. Formal objects, while not observable or measurable in the way that solid objects are, would seem to have as much observer-independent reality as non-formal objects. To take the example of patterns, one might say of a series of numbers that, independently of whether or not any observer exists to notice it, a pattern exists in that series if there is a more efficient

---

[128] Copeland, p.337

way of describing it than an exhaustive description.[131] To take Antony's example of a checkerboard, it seems perfectly reasonable to say that, regardless of what observer is aware of it, the pattern of alternating light and dark squares exists because describing the board in that way is more efficient than describing it square by square. The efficiency is not dependent on observers. Regardless of whether anyone knows it, the pattern of a checkerboard, even a natural checkerboard created by random meteorological phenomena, exists because the efficiency of the potential description exists. If an observer comes along and notices the pattern, the noticing is an act of *discovery*, not one of *creation* or *invention*. In the end, the belief that something must be observer-dependent simply because it is formal seems difficult to uphold. Except possibly for Henri Poincare, who believed that even Euclidean geometry was conventional,[132] few if any philosophers have shared it. Therefore, if Searle's assertion that algorithms or programs do not exist is based on the assumption that 'formal' implies 'observer-relative', then Searle's attack would seem ungrounded.

A second assumption that Searle seems to be making in his claim that things like computations or algorithms do not exist as intrinsic physical features of systems is that algorithms of any sort, regardless of whether they are called procedures, rules, or simply algorithms, cannot exist without being

---

[129] Copeland, p.335
[130] Antony, p.387
[131] Dennett (1991b), p.32 This definition of patterns is originally attributed to Gregory Chaitin, "Randomness and Mathematical Proof", *Scientific American*, CCXXXII (1975), p.48.
[132] Poincare, p.136. I do not mean to make a substantial claim about Poincare,

consciously followed, acknowledged, or thought about in some way by an observer. Searle clearly opposes the idea that an algorithm could be an observer-independent procedure for achieving a specific result because algorithms do not exist unless conscious agents recognize them as such. He explains:

> ... ask yourself what happens when a system implements an algorithm. The human computer consciously goes through the steps of the algorithm, so the process is both causal and logical: logical because the algorithm provides a set of rules for deriving the output symbols from the input symbols, and causal because the agent is making a conscious effort to go through the steps. In the case of the mechanical computer, the whole working system is both causal and logical: logical because the homunculus gives an interpretation to the processes of the machine, and causal because the hardware of the machine causes it to go through the processes. But these conditions cannot be met by the brute, blind, nonconscious neurophysiological operations of the brain. In the brain computer there is no conscious intentional implementation of the algorithm as there is in the human computer, but there can't be any nonconscious implementation as there is in the mechanical computer either, because that requires an outside homunculus to attach a computational interpretation to the physical events.[133]

In my opinion, Searle's claim here is unconvincing. I believe that algorithms can exist independently of anyone's awareness of them, and I see no reason why an algorithm's existence should be essentially linked to consciousness. When Copeland talks about algorithms as lists of machine executable instructions, he leaves the executor of those instructions open. Although a conscious agent can follow the instructions, an unconscious machine can follow them just as well. I can see no inconsistency in saying that the brain might be a computer that implements certain algorithms. Algorithms, quite simply, do not

---

but only to point him out as a possible candidate.

need to be thought about in order to exist. Thus, I can see no reason to accept Searle's claim that 'algorithm' and 'computation' do not name intrinsic physical features of systems, because I can see no reason to accept Searle's assumption that an algorithm cannot exist independently of an observer's awareness of it.[134]

In the end, I believe that Searle has no convincing criticism of the position I propose. Moreover, I believe that, as a general position on the nature of computation and syntax, it is superior to Searle's. To begin with, the two problems that I pointed out earlier in Searle's position have no bearing here. The problems of being intrinsic to physics disappear because I make no claims specifically about being intrinsic to physics. Syntax, as a physical phenomenon, is as intrinsic or extrinsic to physics as any other physical phenomenon is. Just as temperature, voltage, or mass are considered intrinsic properties of the world, so I consider syntax. And just as conventional labelling schemes such as degrees centigrade, volts, and kilograms are observer-dependent, so too are the conventional labelling schemes of specific token syntaxes. I also completely eradicate any slippery slope concerns. Because my position draws a much clearer line between that which is intrinsic to the world and that which is observer-relative,

---

[133] The Rediscovery of the Mind, p.220

[134] Interestingly enough, Searle has a completely different position on the nature of *rule* following. In The Construction of Social Reality, chapter 6, he makes perfectly clear his belief that people can follow rules without being conscious of them and can even follow rules that they have false beliefs about. Clearly, where rules are concerned generally, Searle does not believe that agents need to be conscious of the rules that they follow. Where algorithms are concerned, however, Searle holds the opposite view. As is made clear in the passage that is quoted, Searle clearly believes that algorithms, in order to be

there is no danger of the line being blurred. As a result, my external realism is much more unified than Searle's; whereas he draws a precarious line between things that depend on the thoughts of observers and things that do not, I lump everything on the observer-independent side, and so I have no precarious line to draw. It is not difficult, on my account, to determine where the observer begins to play a crucial role because I do not give the observer a crucial role in the first place. Syntax, just like any other aspect of the physical world, exists independently of our representations of it. Token syntaxes are of course observer-relative, just like any other token labelling scheme, and this is unproblematic. Thus, not only does my position withstand Searle's criticisms of it, but it also avoids the problems inherent in Searle's own position.

Also, I believe my position synchronizes well with the concepts of multiple realizability and natural function, and with the dismissal of any homunculus fallacy. Concerning multiple realizability, since it is the discrete physical phenomena themselves that matter and not the specific labelling scheme that we apply to the phenomena, whether or not a given system is a computer becomes an ordinary empirical question. Given the constraints on interpretation previously discussed, and given the concept of an honest interpretation of a theory, namely Turing's computational theory, it becomes a plain matter of fact whether computation is occurring in a specific situation. Computational syntax is formal in the same way that

---

followed at all, need to be *consciously* followed.

geometrical properties are formal, and are multiply realizable in the way that geometrical phenomena are multiply realizable. Syntax can be realized in more than one way and in more than one substance because it is a formal entity with abstract properties which can be realized in different kinds of physical systems. It is not universally realized because the facts about syntax are observer-independent, and constrained by the restrictions previously outlined. Whether or not a system exhibits a certain syntax is a factual question, even though many systems can exhibit the same syntax. It is also easy to see why, on my theory, there is no homunculus fallacy. Because syntactical phenomena can exist prior to any syntactical labelling scheme, there is no need for a homunculus to fix a labelling scheme onto a physical system. Although the labelling schemes themselves are observer-dependent, the syntactical phenomena that the labelling schemes label are not at all observer-dependent. On my account of syntax, there is simply no need for a homunculus where discrete physical phenomena are concerned, because even in an observerless universe the phenomena would exist all the same. Lastly, my concept of natural syntax is very much in tune with the concept of natural function. Just as functions can exist without observers and can be discovered by observers, so too can syntaxes. Also, just as there is a real sense in which natural functions can be discovered in biology, so too can natural computational syntaxes be discovered within the brain.[135] Natural

[135] Note that I do not make any commitment to cognitivism in this thesis. I am simply clearing the way for the possibility that the brain is a digital computer, and am refuting Searle's assertion that cognitivism is certainly false. Whether or not the brain in fact is a computer does not concern me.

syntax is the perfect compliment to natural function, and a belief in one compliments a belief in the other. Clearly, on the three issues of multiple realizability, natural function, and the disappearance of any homunculus fallacy, my position remains neatly consistent.

There is one last potential problem with my theory that requires dismissal. On my theory, it follows that there is a very large set of possible token syntaxes, or syntactical labelling schemes, that apply to a given syntactical system, and perhaps even an infinite set if pragmatic considerations such as simplicity or brevity are ignored. While this might seem an uncomfortable result to some, in my opinion it is not only unproblematic but also quite natural. When cognitivists, computational theorists, supporters of artificial intelligence, or even philosophers of mind talk about the 1's and 0's of the binary syntax of a desktop computer, I believe all of them understand, at their most honest moments, that the labelling scheme they are using first is only a labelling scheme of an independent syntactical reality and second is only one of many possible labelling schemes that could have been used. I believe that we tend to think of 1's and 0's as literally existing in computers because the labelling scheme we use is maximally simplistic. Since there is no better way of describing the phenomena, we tend to literally identify the description with the thing described. On careful consideration, I believe most of us will readily admit that there are many, perhaps infinite, ways that a given syntactical system such as a desktop could be

described or labelled. The fact that there are many possible labelling schemes for a given syntactical system thus becomes not only unproblematic, but also helpful in reminding us that our particular scheme is not identical with the phenomena being schematized.

Certainly, more work needs to be done in formulating a theory of natural syntax. All that I have offered here is a brief outline, a vague blueprint from which a more precise concept of natural syntax might be formed. However, as brief and sketchy as it is, I believe that my construal of syntax is clearly more plausible and less problematic than Searle's. In my opinion, my account of syntax, despite its vagueness, not only stands up to Searle's but surpasses it. Recall, however, that any superiority over Searle's position is only a bonus. All that is required to over-turn Searle's own unsupported attack on cognitivism, on the claim that the brain is a digital computer, is the existence of at least one other plausible account of syntax and computation. I see no reason whatever to think that my position is not plausible, and so in my opinion the victory is already achieved. The fact that I see every reason to think that my position is superior to and more intuitive than Searle's is only icing on that victory.

## A Final Assessment

What conclusion can be drawn from Searle's attack on cognitivism? In my opinion, there are two very distinct possibilities. On the one hand, one might say that Searle's attack on the possibility that the brain could be a digital

computer is fundamentally flawed because he misunderstands or misrepresents the true nature of syntax. One might assert that Searle's arguments fail because he simply does not have a good understanding of the phenomena involved, and that his position is shown to be inferior when a more accurate account of those phenomena is provided. I do not believe, however, that this is the best possible assessment. Instead, I will say that Searle acts as a valuable litmus paper, illuminating the sloppy understanding of computation that is common in cognitive science, and demonstrating just how erroneous that understanding can be. In my opinion, Searle shows us the danger of confusing syntax as a labelling scheme with syntax as the discrete physical phenomena that we apply a labelling scheme to. Searle helps us see that when we fail to differentiate between the conventional description of phenomena and phenomena themselves, all kinds of problems break out. Searle, in a certain context, is absolutely right; if syntax the labelling scheme is mistaken for syntax the physical phenomenon, and computation is taken to be defined in terms of the labelling syntax, then of course the brain could not be a computer. It is only when syntax is defined as a formal, observer-independent physical phenomenon that it makes sense to say that it is possible for the brain to be a computer. In the end, Searle does cognitivism a service. By pulling on the loose threads of our concepts, he forces us to sharpen our understanding and our use of those concepts, and ultimately to tighten our grasp of the phenomena themselves. In the end, cognitivists come out of it with a tighter knit.

**BIBLIOGRAPHY**

Antony, Louise M. (1997) "Feeling Fine About the Mind", *Philosophy and Phenomenological Research*, 57:381-7.

Bailey, Andrew R. (1994) "Representations versus Regularities: Does Computation Require Representation?", *Eidos* 7:47-58.

Baumgartner, Peter, and Sabine Payr (1995) *Speaking Minds: Interviews with Twenty Eminent Cognitive Scientists*, Princeton: Princeton University Press.

Beards, Andrew (1994) "John Searle and Human Consciousness", *Heythrop* 35:281-295.

Blackburn, Simon (1996) *The Oxford Dictionary of Philosophy*, Oxford: Oxford University Press

Block, Ned (1995) "The Mind as the Software of the Brain", in Daniel N. Osherson, ed., *An Invitation to Cognitive Science*, vol.3: 377-425.

Boorse, C. (1976) "Wright on Functions", *The Philosophical Review*, 85:70-86.

Burton, Robert G. (1995) "Searle on Rediscovering the Mind", *Man and World* 28:163-174.

Chalmers, David (1996) *The Conscious Mind*, New York: Oxford University Press.

Chrisley, Ronald L. (1994) "The Ontological Status of Computational States", in Gianfranco Soldati, ed., *European Review of Philosophy*, vol.1: *Philosophy of Mind*, 55-75.

Churchland, Paul M. (1995) *The Engine of Reason, the Seat of the Soul: A Philosophical Journey Into the Brain*, Cambridge: MIT Press.

Copeland, Jack B. (1993) "The Curious Case of the Chinese Gym", *Synthese* 95:173-186.

Copeland, Jack B. (1996) "What Is Computation?", *Synthese* 108:335-359.

Crane, Tim (1993) rev. of *The Rediscovery of the Mind*, *The International Journal of Philosophical Studies* 1:313-324.

Dennett, Daniel C. (1978) *Brainstorms*, Hassocks: The Harvester Press Limited.

Dennett, Daniel C. (1991a) *Consciousness Explained*, Boston: Little, Brown and Company.

Dennett, Daniel C. (1991b) "Real Patterns", *The Journal of Philosophy*, 88:27-51.

Dennett, Daniel C. (1993) rev. of *The Rediscovery of the Mind*, *The Journal of Philosophy* 90:193-205.

Dennett, Daniel C. (1995) *Darwin's Dangerous Idea*, New York: Touchstone.

Endicott, Ronald P. (1996) "Searle, Syntax, and Observer Relativity", *Canadian Journal of Philosophy* 26(1):101-122.

Fellows, Roger (1995) "Welcome to Wales: Searle on the Computational Theory of Mind", in Roger Fellows, ed., *Philosophy and Technology. (Royal Institute of Philosophy Supplement:38)*.

Fodor, J. A. (1968) "The Appeal to Tacit Knowledge in Psychological Explanation", *The Journal of Philosophy* 65:627-40.

Garrett, Brian J. (1995) "Non-reductionism and John Searle's *The Rediscovery of the Mind*", *Philosophy and Phenomenological Research* 55:209-215.

Honderich, Ted (1995) "Consciousness, Neural Functionalism, Real Subjectivity", *American Philosophical Quarterly* 32:369-381.

Jacob, Pierre (1995) "Consciousness, Intentionality and Function. What Is the Right Order of Explanation?", *Philosophy and Phenomenological Research* 55:195-200.

Kim, Jaegwon (1995) "Mental Causation in Searle's 'Biological Naturalism'", *Philosophy and Phenomenological Research* 55:189-194.

Lepore, Ernest, and Robert Van Gulick (1991) *John Searle and his Critics*, Cambridge: Basil Blackwell.

Lycan, William G. (1981) "Form, Function, and Feel", *Journal of Philosophy* 78:24-50.

Lycan, William G. (1996) "Philosophy of Mind" in *The Blackwell Companion to Philosophy*, Bunnin, N. F. (ed), Cambridge: Blackwell.

MacLennan, Bruce J. (1994) "Words Lie in Our Way", *Minds and Machines* 4:421-437.

Melnyk, Andrew (1996) "Searle's Abstract Argument Against Strong AI", *Synthese* 108:391-419.

Millikan, Ruth G. (1989) "In Defense of Proper Functions", *The Philosophy of Science* 56:288-302.

Nagel, Earnest (1977) "Teleology Revisited", *The Journal of Philosophy* 84:261-301.

Nagel, Thomas (1995) *Other Minds: Critical Essays 1969-1994*, New York: Oxford University Press.

Neander, Karen (1991a) "Functions as Selected Effects: the Conceptual Analyst's Defense", *Philosophy of Science (East Lansing)* 58:168-84.

Neander, Karen (1991b) "The Teleological Notion of 'Function'", *Australasian Journal of Philosophy* 69:454-68.

Olafson, Frederick A. (1994) "Brain Dualism", *Inquiry* 37:253-265.

Poincare, Henri (1952) *Science and Hypothesis*, New York: Dover Publications.

Preston, Beth (1995) "The Ontological Argument Against the Mind-Machine Hypotheses", *Philosophical Studies* 80:131-57.

Putnam, Hilary (1988) *Representation and Reality*, Cambridge: M.I.T. Press.

Schweitzer, Paul (1994) "Intentionality, Qualia, and Mind/Brain Identity", *Minds and Machines* 4:259-282.

Searle, John R. (1980a) "Minds, Brains, and Programs," *Behavioral and Brain Sciences* 3:417-424.

Searle, John R. (1980b) "Intrinsic Intentionality: Reply to Criticisms of 'Minds, Brains, and Programs'", *Behavioral and Brain Sciences* 3:450-456.

Searle, John R. (1982) "The Chinese Room Revisited: Response to Further Commentaries on 'Minds, Brains, and Programs", *Behavioral and Brain Sciences* 5:345-348.

Searle, John R. (1983) *Intentionality: An Essay in the Philosophy of Mind*, Cambridge: Cambridge University Press.

Searle, John R. (1992) *The Rediscovery of the Mind*, Cambridge: MIT Press.

Searle, John (1995a) "Consciousness, the Brain and the Connection Principle: A Reply", *Philosophy and Phenomenological Research* 55:217-232.

Searle, John R. (1995b) *The Construction of Social Reality*, New York: The Free Press.

Searle, John R. (1997) *The Mystery of Consciousness*, New York: New York Review of Books.

Skinner, B. F. "Behaviorism at Fifty", [sound recording], Jeffrey Norton Publishers.

Stoutland, Frederick (1994) "Searle's Consciousness", *Philosophical Books 35:245-254*.

Turing, Alan (1936) "On Computable Numbers, with an Application to the Entscheidungsproblem", *Proceedings of the London Mathematical Society*, Series 2, 42:230-65.

Turing, Alan (1948) "Intelligent Machinery", National Physical Laboratory Report, Reprinted in Meltzer, B. and Michie, D. (eds), *Machine Intelligence*, Edinburgh: Edinburgh University Press, 5:3-23.

Umapathy, Ranjan (1996) "The Mind-Body Problem: A Comparative Study", *Journal of Indian Council Philosophical Research*, 13:25-51.

Van Gulick, Robert (1995) "Why the Connection Argument Doesn't Work", *Philosophy and Phenomenological Research* 55:201-207.

Ward, Andrew (1993) "Searle and the Homuncular Fallacy", *Journal of Indian Council Philosophical Research*, 10:61-72.

Wright, Larry (1973) "Functions", *The Philosophical Review* 82:139-168.

Wright, Larry (1976) *Teleological Explanation*, Berkeley: University of California Press.